

SCIENCE  
LIBRARY

## Psychological Bulletin

## CONTENTS

## ARTICLES:

- Thematic Apperception Test: Interpretive Assumptions and Related Empirical Evidence:* GARDNER LINDZBY, 1.  
*When Not to Factor Analyze:* J. P. GUILFORD, 26.

## NOTES:

- A Note on Christie's: "Experimental Naïveté and Experiential Naïveté":* LEWIS BERNSTEIN, 38.  
*A Note on the Problem of Homogeneity-Heterogeneity in the Use of the Matching Method in Personality Studies:* PAUL F. SECORD, 41.  
*Tests of Hypotheses: One-Sided vs. Two-Sided Alternatives:* LYLE V. JONES, 43.  
*Some Comments on Thistlethwaite's Perception of Latent Learning:* HOWARD H. KENDLER, 47.  
*The Blodgett and Haney Types of Latent Learning Experiment: Reply to Thistlethwaite:* IRVING MALTZMAN, 52.  
*Reply to Kendler and Maltzman:* DONALD THISTLETHWAITE, 61.  
*Correction:* A. RABIN AND W. GUERTIN, 71.

## BOOK REVIEWS:

- BLAKE AND RAMSEY'S *Perception: An approach to personality*: DAVID C. McCLELLAND, 72.  
 FORD AND BEACH'S *Patterns of sexual behavior*: FLIGHT STELLAR, 75.  
 MORGAN AND STELLAR'S *Physiological psychology*: JOSEPH E. BARNACK, 78.  
 BARTLEY'S *Beginning experimental psychology*: CARL PRAETMAN, 80.  
 BUGELSKI'S *A first course in experimental psychology*: DONALD M. JOHNSON, 82.  
 MURPHY'S *An introduction to psychology*: FRANK W. FINGER, 83.  
 GARRETT'S *Psychology*: S. C. ERICKSEN, 85.  
 PRONKO AND BOWLES' *Empirical foundations of psychology*: GEORGE F. J. LEHNER, 86.  
 ARNOLD'S *Das Raumlebens in Naturwissenschaft und Erkenntnistheorie*: KARL F. MUENZINGER, 88.  
 CATTELL'S *Personality: A systematic, theoretical and factual study*: JEAN WALKER MACFARLANE, 89.  
 CATTELL'S *An introduction to personality study*: ALFRED L. BALDWIN, 90.  
 YACOVLEVSKY'S *Medical psychology. A basis for psychiatry and clinical psychology*: SEYMOUR B. SARASON, 92.  
 BARKER AND WRIGHT'S *One boy's day: A specimen record of behavior*: CELIA BURNS STENDLER, 92.

BOOKS AND MATERIALS RECEIVED: 94.

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

LYLE H. LANIER, Editor  
*University of Illinois*

LORRAINE EOUTHILET, Managing Editor

CONSULTING EDITORS

STUART H. BRITT  
*Needham, Louis and Brorby, Inc., Chicago*

DORWIN CARTWRIGHT  
*University of Michigan*

FRANK A. GELDARD  
*University of Virginia*

JAMES J. GIBSON  
*Cornell University*

DAVID A. GRANT  
*University of Wisconsin*

WILLIAM T. HERON  
*University of Minnesota*

ERNEST R. HILGARD  
*Stanford University*

WILLIAM A. HUNT  
*Northwestern University*

JEAN WALKER MARSHALL  
*University of California*

DONALD G. MARSHALL  
*University of California*

JOHN T. METCALF  
*University of Toronto*

JAMES G. MILLER  
*University of Chicago*

NEAL E. MILLER  
*Yale University*

HELEN PRABHU  
*University of Michigan*

ROBERT R. SUGGS  
*Harvard University*

ROBERT L. THORNDIKE  
*Teachers College, Columbia University*

*The Psychological Bulletin* contains evaluative reviews of the literature in various fields of psychology, methodological articles, critical notes, and book reviews. This JOURNAL does not publish reports of original research or original theoretical articles.

Editorial communications, manuscripts and book reviews should be sent to Lyle H. Lanier, Department of Psychology, University of Illinois, Urbana, Illinois.

**Preparation of articles for publication.** Authors are strongly advised to follow the general directions given in the article by Anderson and Valentine, "The preparation of articles for publication in the journals of the American Psychological Association" (*Psychological Bulletin*, 1944, 41, 343-376). Special attention should be given to the section on the preparation of the bibliography (pp. 363-372), since this is a particular source of difficulty in long reviews of research literature. All copy must be double-spaced, including the bibliography.

**Reprints.** Authors may order reprints when returning proof. Five copies of the JOURNAL are supplied gratis to contributors of articles, notes, special reviews, and book reviews. No reprints are supplied gratis.

**Business communications**—including subscriptions, orders of back issues and changes of address—should be sent to the American Psychological Association, 1015 Massachusetts Avenue, N. W., Washington 5, D. C.

Annual subscription: \$7.00 (Foreign \$7.50). Single copies, \$1.25.

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1015 Massachusetts Ave., N.W., Washington 5, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage provided for in Section 538, act of February 25, 1925, authorized August 1, 1951, and related in U.S.A.

Copyright, 1952, by The American Psychological Association, Inc.

# Psychological Bulletin

## THEMATIC APPERCEPTION TEST: INTERPRETIVE ASSUMPTIONS AND RELATED EMPIRICAL EVIDENCE<sup>1</sup>

GARDNER LINDZEY

*Harvard University*

The chief purpose of this article is to state the assumptions customarily involved in interpreting the Thematic Apperception Test,<sup>2</sup> and to examine the logical considerations and some of the empirical evidence that can be used to verify or reject each of these assumptions.

Aside from certain historical ties to psychoanalysis, the theoretical and empirical continuity between projective testing and the remainder of psychology has been a subject of little interest. This is true in spite of an increasingly large amount of research and formulation in other areas of psychology that is directly pertinent to the activities involved in projective testing. In particular, the research of the "new-look" perceptionists represents an important and fertile link between projective testing and more traditional domains of psychology. The initial outline of such a continuity has been traced by Blake and Wilson (7) in a study of the influence of depressive tendencies upon selectivity in Rorschach response, Bruner (9) in a discussion of the Rorschach test, Lawrence (39) in an investigation of temporal factors in perception, Siipola (75) in an examination of the effect of color upon Rorschach response and Stein (76) in an ingenious study employing tachistoscopic exposure of Rorschach stimulus material. All of these studies illustrate the feasibility and fruitfulness of relating perception findings to material elicited by Rorschach-like techniques. One of the aims of the present paper is to stress the desirability of relating such theory and research to the findings of investigators who employ the Thematic Apperception Test.

<sup>1</sup> This paper is an outgrowth of a study of personality and the imaginative processes directed by Henry A. Murray and supported by grants from Rockefeller Foundation and the Laboratory of Social Relations, Harvard University. I am deeply grateful to a number of colleagues for their generosity in reading and criticizing this manuscript. My greatest single debt is to Henry A. Murray whose constant stimulation and encouragement made this article possible.

<sup>2</sup> These same assumptions are customarily employed in connection with other story-construction projective techniques, e.g., Make-A-Picture-Story Test, Four-Picture Test, Tri-Dimensional Apperception Test, etc.

One may legitimately object to my proposal to appraise empirically statements that I am treating as given, or axiomatic. However, it is not clear that all psychologists would concur in giving axiomatic status to these statements. As is true of so much of psychological formulation, the distinction here between analytic and empirical is not clearly delineated—what is one person's empirical generalization is another's axiom. Further, I believe most psychologists agree with MacCorquodale and Meehl (44) in subscribing to a methodological position that emphasizes the use of *only* analytic constructs that interact smoothly with available empirical knowledge. Consequently, examination of these assumptions in the light of empirical evidence seems justified on the one hand by the questionable axiomatic status of the assumptions and on the other by the general acceptance of psychologists that axioms should not violate observational data.

The assumptions to follow vary greatly in their generality. In fact, they are logically related only by virtue of their frequent use in the interpretation of material secured from one particular type of projective technique. They do not represent all of the assumptions employed in making such interpretations. I have attempted to include only those that are not sufficiently general to be common among all psychologists and yet, within the prescribed situation, are sufficiently common so that almost any individual engaged in this activity would employ them. For the latter reason, I am not concerned here with the assumptions involved in "formal" or "sign" analysis of story projective material where the meaning, or the thematic qualities, of the material is minimized, e.g., Balken and Masserman (3) and Wyatt (84). Nor have I attempted to explore the assumptions involved in the more recent "sequence analysis" (1) of projective responses.

#### THE ASSUMPTIONS AND RELATED EVIDENCE

The assumptions are divided into three crude groups. First, the most general assumption, fundamental to all projective testing. Second, those assumptions that are concerned with procedures employed in determining the diagnostically significant portions of the fantasy productions. Third, those assumptions involved in relating the significant portions of the protocols to other forms of behavior.

*Primary Assumption:* In completing or structuring an incomplete or unstructured situation, the individual may reveal his own strivings, dispositions, and conflicts.<sup>3</sup>

<sup>3</sup> The terms "strivings, dispositions, and conflicts" are meant to designate all the attributes or aspects of the person that the clinician is interested in or wishes to measure. One could readily add to this list such terms as: personality organization, primitive fixations, complexes, needs and press, or any others that seemed necessary to represent those aspects of the person that are being explored.

not concerned  
with formal  
analysis

fundamental  
to all projective  
testing



*Assumptions Involved in Determining Revealing Portions of Stories:*

1. In the process of creating a story the story-teller ordinarily identifies with one person<sup>4</sup> in the drama, and the wishes, strivings, and conflicts of this imaginary person may reflect those of the story-teller.

a. It is assumed further that the identification figure can be established through the application of a number of specific criteria; e.g., person appearing first in the story, person doing most of the behaving, person most similar to story-teller, etc.

b. It is also assumed that additional figures in the stories such as father, mother, or brother often may be equated to the real-life counterparts of the story-teller and the behavior of the hero toward them used as indicative of the story-teller's reactions to these persons.

2. The story-teller's dispositions, strivings, and conflicts are sometimes represented indirectly or symbolically.

3. All of the stories that the subject creates are not of equal importance as diagnostic of his impulses and conflicts. Certain crucial stories may provide a very large amount of valid diagnostic material while others may supply little or none.

4. Themes or story-elements that appear to have arisen directly out of the stimulus material are less apt to be significant than those that do not appear to have been directly determined by the stimulus material.

5. Themes that are recurrent in a series of stories are particularly apt to mirror the impulses and conflicts of the story-teller.

*Assumptions Involved in Deriving From Revealing Portions of Fantasy Material Inferences About Other Aspects of Behavior:*

1. The stories may reflect not only the enduring dispositions and conflicts of the subject, but also conflicts and impulses that are momentarily aroused by some force in the immediate present.

a. The further assumption is frequently made that both the enduring and temporary processes are reflected in stories in the same manner.

2. The stories may reflect events from the past of the subject that he has not himself actively experienced, but rather has witnessed or observed, e.g., street scene, story, motion picture.

a. It is assumed further that, although the subject has not himself experienced these events, and is telling them as he observed them, the fact that he selects these events, rather than others, is in itself indicative of his own impulses and conflicts.

3. The stories may reflect group-membership or socio-cultural determinants in addition to individual or personal determinants.

4. The dispositions and conflicts that may be inferred from the story-teller's creations are not always reflected directly in overt behavior or consciousness.

I wish to emphasize that the research to be discussed in connection with the various assumptions can *not* serve as a validation of this kind of projective testing. At most, it can demonstrate that the very general assumptions lying behind this kind of activity are not in direct conflict

<sup>4</sup> Most investigators accept the possibility of multiple identifications. However, this necessitates the same assumption and only complicates somewhat the general problems of establishing interpretive rules.

with available empirical findings. This may increase the "plausibility" of projective interpretations, but the task of demonstrating the utility of specific rules of interpretation remains.

It is likewise important to realize that in this article I am omitting from consideration most of the pertinent empirical evidence resulting from clinical use of this technique. The omission of this idiographic research or observation is not the consequence of any feeling that such data are not of immense value. Rather, in an article emphasizing the relation of projective testing to the remainder of psychology, it seemed desirable to stress most heavily those kinds of research that came closest to meeting the experimenter's demand for empirical control and intersubjectivity of method. Thus, most of the evidence to be referred to has been secured with some attempt at maintaining adequate empirical controls.

*Does the Individual Reveal His Own Dispositions and Conflicts in Completing an Unstructured Situation?*

This assumption is not limited to story-construction tests but lies at the heart of all projective testing. Fortunately, in view of its ubiquity, there is a host of very general experimental verification, some of it drawn from laboratory research.

One of the first investigators to point to the relationship between motivational factors and response in an unstructured situation was Sanford (71, 72). He demonstrated in a series of experiments that the food responses of subjects varied as a function of the amount of food deprivation they had undergone. The responses were given in the act of telling stories to ambiguous pictures, making word associations, and in other situations where the stimulus was sufficiently unstructured to permit either food or non-food responses. This same general function was later demonstrated under somewhat different conditions by Levine, Chein, and Murphy (40) and McClelland and Atkinson (48). The results of these three investigations do not agree in detail, but all were able to demonstrate some kind of variation in response to ambiguous stimulus situations as a function of food deprivation.

In an early study by Murray (55), it was shown that estimates of "maliciousness" of faces in photographs varied directly with an experimentally induced state of fear. After a fear-producing game children rated a series of photographed faces as significantly more malicious than they had rated the faces previous to the game. A recent study by Katz (33) has shown that the manner in which an individual completes an incompletely drawn face varies systematically with the kind of experience he has undergone previous to the activity. Thus, a group that had just failed on a test of reasoning ability differed significantly from a group that had just spent their time in a neutral activity. Presumably

the manner in which the individual completed this unstructured situation was directly related to the motivational or emotional conditions aroused by the experimental treatment.

A series of investigations by Bruner and Postman (12) have demonstrated that when stimulus material is made ambiguous by very brief tachistoscopic exposure, the responses or "guesses" of the subjects vary systematically with motivational variables. These same investigators, as a result of their perception research, have come to occupy a much more extreme position than that implied by the above assumption. They suggest that:

Most experimenters who have worked with need and attitude factors in perception have assumed, sometimes quite explicitly, that only in highly equivocal stimulus situations can such "nonsensory" factors operate. . . . But all stimulus situations are potentially equivocal and cease to be so only to the extent that selection, accentuation, and fixation have taken place. Perception occurring without the contribution of such adaptive factors is as unthinkable as perception without the mediation of receptive nerve tissue (11, p. 301). . . . *Adaptive factors in perception are not limited to unstable stimulus situations* (11, p. 307).

In the light of this evidence, the assumption that motivational factors are revealed in completing unstructured situations seems clearly warranted.

*Does the Subject Identify with Some Figure in the Stories He Tells  
and Can This Figure Be Established Reliably?*

There seem to be three modal positions that can be maintained in regard to identification in the story-telling process. First, we may make the assumption, indicated above, that there is ordinarily a single identification figure in each story. This assumption can be complicated greatly by a number of special conditions of the type already suggested by Murray (56). Second, we may assume a continuum of identification with those figures in the stories that are very similar to the story-teller possessing a maximum of the subject's attributes and those that are very dissimilar possessing a minimum. This is the assumption made by Sears (74), who, working primarily with doll techniques and heavily influenced by the notion of "stimulus generalization," has suggested specifiable dimensions along which the degree of identification can be expected to vary. Thus, in doll play figures resembling the subject in age and sex will show more characteristics of the subject than those figures that are dissimilar in age and sex. Third, we might, with Henry (30), Piotrowski (58), and others, make no attempt to locate the hero but simply look on all characters in the constructed stories as representative of aspects of the story-teller. This last alternative is perhaps the least happy, not only because a very large amount of diverse clinical experience militates against it, but also because it leads to certain drastic limitations upon the diagnostic use of the TAT. If

we adopt this assumption, we are more or less forced to give up the attempt to appraise the subject's attitudes toward other persons. Thus, Piotrowski suggests that we eliminate the hero and assume "that every figure in the TAT stories expresses some aspect of the testee's personality" (58, p. 107), while somewhat later he suggests that many stories "reflect what the subject thinks and feels about persons represented by the TAT figures, i.e., about the old and the young, the male and the female" (58, p. 113). Thus, with no rules for differentiating, we are told that all figures represent characteristics of the hero, but also that the characteristics of some figures represent the story-teller's attitudes toward other persons. If we assume that all figures are equally representative of the story-teller's characteristics, either we must give up attempts to appraise the subject's attitudes toward other persons through this instrument or else engage in some kind of dialectic in the effort to defend such attempts.

There is, as yet, no unequivocal answer to the question of which of the above assumptions is most useful. In order to show that the assumption of a single identification figure is warranted, it is necessary to demonstrate relationships between the behavior of the subject and the imaginary behavior of the hero that do not exist between the behavior of the subject and the behavior of non-hero figures. If this difference is not shown, one can always suggest that the sensitivity of the test to aspects of the subject is a result not of any specific identification process, but rather of a generalized reflection of motivational state of the sort that Murray (55), Bruner and Postman (12), and others have reported. Thus, even if all figures in the stories represented aspects of the story-teller, we might arbitrarily decide to use only a certain percentage of these figures as diagnostic of the story-teller's attributes, i.e., employ the hero and non-hero distinction, and we would still expect the test to show some sensitivity to variations in story-teller behavior even though we were wasting much of its power. Just as the subject's motivational states are reflected in free association, or in the pre-solution hypotheses of tachistoscopic response, so also the verbal flow accompanying the story-telling process may mirror motivational states without any intervening identification with particular actors in the stories.

The distinction between "heroes" or identification figures on the one hand and "non-heroes" on the other is greatly complicated by the fact that the story-teller's orientation or attitudes toward other persons is intimately related to his own psychic makeup. Thus, we might mistakenly employ an identification figure as representative of the story-teller's attitudes toward other people and actually secure considerable information about the individual's external orientation because of the similarity between "self" and "perceived other." Or, conversely, we might take a figure intended by the story-teller as an "other" object



and find mirrored in it much of real pertinence to the story-teller's own personality structure.

What is needed to answer the above question is a clear demonstration of more intimate relations between "hero" behavior or attributes and the story-teller than can be shown between "non-hero" behavior and the story-teller. Results such as Bellak's (5) showing that the frequency of aggressive words rises when the individual is frustrated do not demonstrate that the identification figure is necessarily displaying more aggressive behavior. Nor are studies satisfactory that simply demonstrate variations in "hero" activities or attributes that relate to behavior or attributes of the story-teller. Even if Bellak had shown that aggressive behavior on the part of heroes rose following frustration, this would still not provide the needed information. In this case it is quite possible that analysis of the behavior of the non-identification figures would reveal the same increase in aggressive behavior, indicating not identification, but simply an increase in verbal responses pertinent to the need or conflict in question.

Lindzey (43) has demonstrated that following a social frustration the incidence of aggressive acts in TAT stories carried out by heroes against others increased more than the incidence of aggressive acts carried out by "other" figures. If the incidence of the two kinds of acts is combined, the resulting shift is greater than the shift in either "self" or "other" figures alone. These findings are complicated by the fact that the frustration situation was of such a nature that both extra-punitive responses and a view of the environment as hostile and threatening would be expected to rise following the experimental treatment. Thus, the increase in aggressive responses on the part of non-hero figures might be considered a result of the increase in aggressive tendencies on the part of this story-teller or a reflection of the fact that the story-teller viewed the external world as a more hostile and threatening place. It is clear that what is needed is an experimental treatment where, given the subject's identification with a hero-figure, the predictions to be made for "hero" and "other" figures will either be opposed or widely different.

Accepting the desirability of locating identification figures, can these figures be specified precisely? Several studies (29, 79) in which relatively high inter-scorer reliability coefficients were secured imply that it is possible to obtain reliable agreement among different scorers in establishing the identification figure. Presumably, a positive correlation for need and press ratings could be secured only if the different raters were treating the same figures as hero. In addition, Mayman and Kutner (45) report complete agreement in 89 per cent of the cases between two raters who independently determined the identification figure for a series of 91 stories.

In general, then, the feasibility of the identification assumption



cannot be clearly demonstrated at present, although empirical evidence suggests that identification figures can be established with reasonable reliability.

*Are Impulses and Conflicts of the Subject Sometimes Represented Symbolically?*

The clinical use of the concept of symbolism is intimately related to a number of psychological concepts employed by more rigorous and empirically oriented investigators. All behavior theorists have encountered the thorny problems posed by interchangeability or substitutability in behavior of stimuli and responses. These problems have led to the formulation of a number of psychological concepts including displacement, substitution, stimulus generalization, and vicarious mediation. Apparently most psychologists agree that when a response is interfered with either by internal or external barriers it will frequently become altered to another form or else be directed toward a new object. There is a large body of research, much of it carried out on animal subjects, demonstrating this relationship and exploring some of the conditions under which it operates (8, 15, 25, 31, 32, 38, 62, 63, 83). Of the animal investigations, perhaps the most pertinent to our discussion is Miller's (54) demonstration that a learned aggressive response can be generalized or displaced from an original stimulus object (another rat) to a substitute stimulus (doll) when the first stimulus is made unavailable. This process can be considered a rough paradigm of symbolic representation.

- 1.
2. A special question of considerable importance to the projective tester is whether symbolic transformations can, and customarily do, take place without awareness of the subject. Most of the clinical studies discussed below deal with symbolic representations which the subject was not aware of initially and which he could become aware of only under special conditions, e.g., interpretation or therapeutic change. This capacity of the organism to engage in symbolic representation without awareness of the process has been demonstrated under better controlled circumstances by other investigators. Diven (17) and Haggard (26) have both shown that a word may become a substitute stimulus for an electric shock and evoke a galvanic skin response that differentiates it from "non-shock" words, even where the subject is unaware of the connection between the word and the electric shock. In similar fashion, McCleary and Lazarus (46) showed that words to which a galvanic skin response had been conditioned through the use of electric shock could elicit the galvanic skin response even when exposed tachistoscopically at such rapid speeds as to prevent conscious recognition.

Certainly since Freud, the clinical interpretation of diagnostic material, whether projective or not, has depended heavily upon the assumption of symbolic transformations. There is some clinical-experimental and a host of clinical-observational evidence to indicate that

symbolic transformations are quite customary in human behavior, particularly where unacceptable or antisocial impulses are involved. ↙

In case histories conversion symptoms and compulsions frequently present themselves dramatically as evidence of the fact that a given conflict or forbidden impulse may secure expression in a manner only indirectly related to the original impulse. Thus, it is generally accepted that hand-washing compulsions often represent symbolically the desire of the individual to cleanse himself of the consequences of masturbation. Likewise, in doll therapy there is excellent evidence that the child is able to discharge against surrogate objects the same impulses and feelings that he has developed toward persons in the real world. The extent to which this relationship between symbol and real-life counterpart is unequivocally demonstrated varies with the individual case. However, in many instances the relationship is effectively demonstrated as the symbolic behavior can be shown to vary directly with changes in the relationship between the subject and the referent of the symbol. Thus, the aggression directed against the father-doll may show a high inverse relationship to the extent to which the child is able to express such impulses against the original stimulus object—the father. Similarly, a symptom may disappear when the impulse or conflict that it is presumed to represent is dissolved by therapeutic procedure or through changes in the balance of environmental determinants. Levy (42) has described a number of clinical cases involving relatively convincing demonstration of symbolic representation of sibling rivalry, fear of castration, defecation, etc. Tomkins (80) reports the study of a single case under varying degrees of drunkenness. He found that in certain cases the hypothesized referent of the symbol was more and more directly represented as the amount of alcohol consumed increased.

Several investigators have attempted to combine the subject matter of clinical psychology with attempts at empirical control in the study of symbolic representation. An interesting study by Farber and Fisher (21) attempted to demonstrate the process of symbolic transformation by means of using hypnosis and asking subjects to create and interpret dreams. Although the study is not well controlled, if we accept the naïveté of the subjects, the evidence is impressive, for some of the subjects, that not only is symbolic representation an observed form of behavior, but also that the process of assigning symbols in large part corresponds with the rules established by Freud from his examination of dream protocols. Consistently, Krout (37) in a better-controlled investigation studied interpretations based upon the assumption that individuals responded to line drawings, representative of basic experiences or objects, *as if* they were the symbol referent. She found that these interpretations could be verified through the inspection of independently collected validation data. In other words, the subjects did appear to respond to the line drawings in a manner consistent with the nature of the object that the Freudian would view as lying behind

this symbol. Franck (22) and Franck and Rosen (23) have reported two studies in which there is some evidence that the response of the individual to male and female symbols varies with his sex and relative maturity. Klein (34), employing the method of hypnotically inducing dreams, was able to show that the dreamer characteristically transformed or disguised external stimuli that were incorporated into the dream.

In general, then, controlled empirical evidence supplements clinical observation to imply the existence of tendencies for human subjects to represent dispositions and conflicts symbolically. The rules for determining the symbol-referent relations are as yet imperfectly understood, although there is evidence indicating the Freudian view of the symbolic process possesses some utility.

*Are the Stories That the Subject Creates of Unequal Importance as Diagnostic of His Impulses and Conflicts?*

It is possible to assume that any behavioral information, no matter how scanty, contains in it the necessary elements to permit a complete understanding of the individual in question. This derivative of strict Freudian determinism is still defended by some investigators. It implies a theoretical model that is able to incorporate all possible empirical relationships and predict these relationships so precisely that once a single value has been inserted into the closed theoretical system, every other construct in the system can be assigned a value. In view of the present state of psychological theory, this assumption seems an exceedingly unwise one. Nor does it appear to guide the actual behavior of most clinicians. The tendency to view different stories as possessing different degrees of psychological significance is rather clearly revealed by the fact that not infrequently, even with twenty stories, the investigator will feel he has not sufficient material to make a diagnosis or come to any understanding of the dynamics of the individual in question. Apparently just as some sets of twenty stories are more difficult to interpret and are less revealing, so also some individual stories are less rewarding than others.

Although presumably implicitly present in the minds of most interpreters, the assumption is seldom mentioned and there has been little systematic attempt to state any criteria by means of which the more important stories are to be separated from the less important. Rapaport (61), who represents an exception to this generalization, has suggested that the more important stories may be distinguished in terms of certain "formal variables." He cites as illustration of these distinguishing characteristics: the consistency of the story with other stories of this individual, its consistency with stories told by other persons to this picture, the faithfulness with which the subject followed instructions in telling this story, and, finally, the extent to which the individual

"has perceived and apperceived the picture adequately in all its parts."

Indirect evidence bearing on this question is provided by the common observation that stories derived from particular cards frequently have little information pertinent to given variables. Thus the individual interested in studying aggression may find that stories told to certain TAT cards will only rarely provide pertinent evidence. Not only do the stories told by the subject vary in significance but *for given purposes* the stories characteristically evoked by different cards vary in significance. Eron (19, 20) has shown that both incidence of themes and the emotional tone of stories differ significantly among the various TAT cards.

Although practical considerations and clinical experience indicate the necessity of an assumption of "crucial stories," there seems to be little empirical evidence that is directly pertinent.

*Are Themes or Story Elements That Appear to Arise Directly out of the Stimulus Material Less Significant than Those That Are Not Directly Determined by the Stimulus Material?*

The extent to which this assumption is embraced by projective testers is made clear in the general emphasis upon the importance of bizarre responses, sex reversals, etc. Rotter (69) gives it explicit stress in his interpretive discussion of the importance of "unusualness" of response in the Thematic Apperception Test. In the scoring of sentence completion responses, Rotter (70) has reported finding special diagnostic significance associated with "twist" or "reversal" endings to incomplete sentences. Weisskopf (81) has emphasized the assumption heavily in her suggestion that one method of estimating the diagnostic efficiency of projective stimulus material is through the use of a "transcendence index." This index is derived from the descriptive comments of the subject in responding to the stimulus material that go beyond "pure description."

The first difficulty we encounter in attempting to appraise this assumption is the question of how to go about measuring the degree to which a given response is determined by the stimulus material. This could be determined by estimating the structural similarities between the response and the stimulus. Or, one might develop empirical norms representing the common or usual response elements for each stimulus. Rosenzweig (67, 68) has chosen the second alternative in his attempt to establish "apperceptive norms." The function of these norms, in large part, is to enable the interpreter to differentiate that which derives naturally from the stimulus card from that which is projective or personally determined. The norms reported by Eron (19) could be used in similar fashion. Presumably most clinicians implicitly employ both of these approaches. On the one hand they note responses which are an obvious denial of the observable elements in the stimulus materi-



al, and on the other they build certain expectations in regard to what a "normal" story is to each stimulus and they tend to place special interpretive emphasis upon departures from these norms.

Some projective testers probably object to this assumption on the grounds that "any" response, no matter how normal or directly derived from the stimulus material, may have significance in projective protocols. They might point to this attribute as one of the chief differentiae between mental testing (normative) and projective testing (idiosyncratic). It is certainly true that the extent to which a response deviates from the norm is much less a matter of concern to the interpreter of projective protocols than it is to the "mental tester." However, the fact remains that if a particular response is given by one hundred per cent of the persons taking the test, this response cannot possess diagnostic significance unless it is combined with some other response element that is less frequently encountered. In the latter case it is clear that the second or less frequent response is determining the interpretation, not the "normal" response. Nevertheless, it must be remembered that even a response given by a very high percentage of the respondents may have broad significance as in the case of a yes-no response to a specific question. In the limiting case, a response may be given by ninety-nine per cent of the respondents and still be an important diagnostic sign, especially in those few cases where it is not encountered.

Limited empirical justification for the view that significance and "stimulus-boundedness" are negatively related is provided in a study by McClelland, Burney, and Roby (49). They observed that *introduction* by the story-teller of an affiliated person into TAT stories where this person was not present in the picture was related to an experience the subjects had just undergone. A count of affiliated persons without consideration of whether or not they were in the picture did not reveal any relationship with the experimental treatment.

Supporting this assumption, we have the observation that most perception researchers accept the generalization that the more ambiguous the material, the easier it is to observe the operation of directive or motivational factors. Thus, they suggest that in those cases where the response bears relatively little relation to the stimulus material we are especially apt to observe dynamic or motivational factors. Closely related also is Bruner's (10) and Postman's (59) suggestion that the perceptual process may be represented as involving a relationship between the predetermining tendencies of the individual on the one hand, and information or stimulus constraints presented by the environment on the other. They imply that the stronger the determining tendency, the less in the way of environmental supports are needed to produce a related percept and the more contradictory information will be needed to produce a percept in opposition to the predetermining tendency. The above assumption is quite similar to Bruner's and Postman's



view, since it proposes that the stronger the stimulus constraints or supports, the less we know about the predetermining tendencies in the individual that have produced a response consistent with these constraints. However, if the stimulus material is not consistent with the report of the subject, presumably the predetermining tendency was sufficiently strong to produce an appropriate response, even though the stimulus did not call for it.

Thus, the assumption that stimulus-bound responses are less diagnostic than responses that do not depend so heavily upon stimulus constraints appears to fit well with available empirical data.

*Are Recurrent Themes in a Series of Stories Particularly Likely to Mirror the Impulses and Conflicts of the Story-Teller?*

This is an assumption frequently stated but for which there is a minimum of pertinent empirical evidence. However, there are heavy rational considerations favoring such a view. The presence of the same theme even when the stimulus situation has been thoroughly altered implies strongly that there are impelling forces within the individual creating these themes rather than their being the inevitable outcome of stimulus constraints. The same theme following a change in stimulus material suggests that the response is not tied directly to, or produced by, a single specific stimulus. Thus, we may infer that the response tendency possesses generality and is more likely to be related to motivational factors than the response linked to a single stimulus. Even if we accept the view that in all instances the stimulus is evoking the response at question, we know that one measure of the strength of a drive is the extent to which stimulus generalization or displacement will occur. Thus, if the individual can equate a large number of stimuli in order to make this response, we may infer that the instigation lying behind the response is quite strong.

In addition, if interpretations are favored that incorporate a large amount of material, those based upon recurrent themes have a natural advantage in that the interpretation of one theme is automatically applicable to all of the other themes in the series. Finally, the presence of recurrent themes permits the investigator to sample or test the conditions under which these themes make their appearance.

In general, then, this assumption seems to have been accepted consistently by most projective testers although there is little or no direct evidence demonstrating its utility. However, there are a number of rational considerations that support such an assumption.

*Do Stories Reflect not only Enduring Dispositions but also Momentary Impulses and Are These Both Reflected in the Same Manner?*

Most investigators are primarily interested in the Thematic Apperception Test as a means of measuring enduring dispositions, although

Relating Story  
to Form of  
Personality  
T20  
fours  
Ela

occasionally, especially in research, the measurement of situational or temporary instigations to behavior may be crucial.

There have been numerous studies showing the sensitivity of the instrument to temporary or situational determinants. Bellak (5) showed that the number of aggressive words in TAT stories increased when the story-teller was rebuked for the low quality of the stories he told. Sanford (71, 72) and Atkinson and McClelland (2) have shown that TAT protocols vary with food deprivation, and McClelland, Clark, and Roby (50) have shown story variation as a result of exposure to failure in a test situation. Rodnick and Klebanoff (66) have shown that TAT stories vary with relative success in a level of aspiration test. Lindzey (43) demonstrated that extrapunitive behavior on the part of the hero in TAT protocols increased significantly following failure in a social situation.

In similar fashion there are a number of studies that show the sensitivity of the stories to more enduring dispositions. It seems reasonable to accept individuals belonging to different psychiatric groups as differing in some enduring rather than situational attribute. Consequently, studies showing significant differences between groups separated on some diagnostic variable may be considered evidence of the test's sensitivity to enduring dispositions and conflicts. Balken and Masserman (3) observed significant differences in the TAT performance of patients categorized as conversion hysterics, anxiety hysterics, and obsessive compulsives. Renaud (64) was able with some difficulty to distinguish between psychoneurotics, traumatic brain disorder cases, and brain disease cases on the basis of TAT protocols. Cox and Sargent (16) found differentiating signs in the TAT performance of "stable" and "disturbed" school children. Working with mental hospital inmates Harrison (27) in approximately 77 per cent of the cases was able to identify accurately on the basis of TAT stories the psychiatric category in which the patient had been placed. He reports (28) similar results when "blind analysis" was employed, that is, when the tests were administered by a different person than the interpreter.

Further evidence of the sensitivity of the test to non-situationally determined motivation is supplied by studies such as Murray and Stein's (57) where a relatively high positive relationship was found between leadership ratings of ROTC candidates based on TAT performance and leadership ratings independently executed by officers of the men in question. Likewise, Henry's (30) study of the Navaho and Hopi suggests the ability of the test to discriminate these two groups, and further to supply psychological information concerning them that is consistent with information secured from extended observation or through the use of independent instruments. White (82) showed that there was a significant relationship between TAT response and the dispositions leading to hypnotizability. Harrison (27) in those cases where he was able to match descriptive statements based upon TAT responses with

independent information derived from mental hospital case histories was correct in 82.5 per cent of the instances.

The question of whether temporary and enduring tendencies are reflected in stories in exactly the same fashion is important for two reasons. First, the clinician is primarily concerned with the more enduring tendencies and, therefore, it is desirable that he have some means of differentiating between these two classes of determinants. Second, some research, especially McClelland's (2, 50) has implied that through studying the effect of situational factors upon TAT performance it is possible to arrive at means of interpreting TAT stories to reveal the operation of more enduring tendencies. If the reflection of these two kinds of motivational factors should prove to be very similar or the same process this would be unfortunate for the clinician, who would then be faced with the difficulty or impossibility of knowing whether a given tendency was temporarily instigated or whether this was a more permanent characteristic of the individual in question. On the other hand, such similarity in process would encourage experimental treatments as feasible means of approaching the task of specifying more exactly the means of inferring enduring tendencies.

McClelland (47) has supplied some evidence that the enduring and temporary processes are reflected in the same manner. He has demonstrated that the differences in subjects' TAT responses following experimental induction of a motivational state (threatening test situation) is related to other measures of the individuals' behavior, e.g., academic performance. This implies that the response of the subject to the immediate situation is related to his more permanent patterns of response.

McClelland and Liberman (51) derived a system for scoring need achievement in TAT responses that was based upon the differences produced in TAT protocols following experimentally induced failure in a situation related to achievement. In addition, they demonstrated that achievement as measured by this scoring technique was related to performance in an anagrams test and also related to the speed with which certain kinds of achievement related words could be recognized when exposed tachistoscopically. Thus, a measure of need achievement derived from a temporary instigation appeared to be related to other measures of achievement which were only distantly related to the initial instigating situation. While this evidence is not compelling, it does provide some support for the notion that temporary instigations affect TAT stories in a manner consistent with the way in which more enduring dispositions influence stories.

We appear to have excellent empirical evidence indicating that stories are responsive to both situational and enduring motivational factors. There is no conclusive evidence demonstrating the similarity or dissimilarity of the process whereby these two classes of determinants secure expression in the stories.

*Do the Stories Reflect Events from the Past of the Subject That He Has Not Himself Actively Experienced? Are These Events Diagnostic of the Individual's Dispositions and Conflicts?*

yes  
Inquiry following the customary administration of the TAT has verified the hypothesis that individuals do incorporate material taken directly from scenes that they have witnessed or from movies or books they have been exposed to. Murray (56) has reported this. Further, any clinician who has worked at all extensively with the TAT has inevitably many cases in his own experience of stories drawn directly from the world of the novel or drama, frequently with accompanying remarks indicating explicitly that this was the case. This is directly consistent with Freud's (24) early dictum that each dream incorporates something from the events of the preceding day.

Given the influence of these non-participated events, the question then becomes one of whether motivational factors affect or are revealed in the recall of such experiences. The selective function of memory and the importance of motivational determinants in the memory process have been recognized for some time, certainly since the appearance of Bartlett's (4) treatise on memory. A number of studies have investigated, under reasonably well-controlled conditions, memory as a function of such motivational variables as political ideology, Edwards (18), Levine and Murphy (41); value as measured by the *Study of Values*, McGinnies and Bowles (52); sex membership, Clark (14); mental set, Carmichael, Hogan and Walter (13); attitude, Postman and Murphy (60); punishment, McGranahan (53). Although there are many other factors, e.g., primacy, recency, vividness, known to influence recall, it seems reasonably well established that motivational factors do serve as one important class of determinants of memory.

If we accept these studies as evidence of the extent to which memory is influenced by motivational factors, it seems reasonable that the individual, in the process of recalling past events or experiences, will reveal or expose important aspects of his own motivational state. Consequently, the assumption that the particular events remembered by the subject are diagnostic of his dispositions and conflicts appears to be supported by available empirical data.

*Do Stories Reflect Group-Membership, Cultural or Social Determinants as well as Personal or Individual Determinants?*

This assumption simply implies consistent differences between the fantasy productions of individuals who belong to, or have been socialized in, different social groups. Thus, a certain amount of the variation in any TAT production can be accounted for by the fact that the individual has grown up in a given milieu or social role. The importance of the assumption derives from the fact that overlooking this kind of variation introduces a serious source of error in the interpretation of



imaginative protocols from members of more than one social group.

Although little effort has been made to explore the variations in fantasy productions between many of the important groups of our own society, it is quite widely accepted or expected that these differences exist. Even such an important cleavage as that between male and female has been little explored so far as TAT behavior is concerned, while such variables as socio-economic status, occupational role, and ethnic group-membership have also been of slight interest to most investigators. Rosenzweig and Fleming (68) have reported differences between a roughly equated group of men and women on a number of specific aspects of TAT response. An investigation by Riess, Schwartz, and Cottingham (65) designed as a critical appraisal of the Thompson (78) modification of the TAT led to the observation of certain relatively slight variations in story length as a function of geographic residence and Negro-white group-membership.

Henry (30) in his investigation of Navaho and Hopi children found that inferences based upon his adaptation of the TAT related to independently secured information concerning the children and also that there were systematic differences between the Navaho and Hopi fantasy productions. In addition, he reports differences between Navaho subjects who were members of different subgroups in this society.

Although there is a paucity of empirical evidence demonstrating differences in fantasy production between various socio-cultural groups, what evidence is available appears to support this assumption.

*Are Impulses and Conflicts Inferred from Stories Not Always Reflected Directly in Overt Behavior and Consciousness?*

Projective testers vary considerably in the manner in which they emphasize this assumption. Some see a very intimate relation between imaginative behavior and overt behavior. Thus, Piotrowski introduces nine rules designed to permit the translation of fantasied into overt with the following statement:

The rules proposed in this article are a new attempt to solve the problem of the relationship between the TAT and overt behavior. Since the TAT is mainly an exercise in creative imagination, it should reflect the patient's ideas and drives regardless of whether or not they find a direct expression in overt behavior. Thus, parts of the TAT always reflect the overt behavior of the subject while other parts reflect ideas which are not as directly manifested in overt actions. If this be so, we need a rule by means of which we could differentiate these two parts of the TAT. The rules presented below have been formulated largely for the purpose of meeting that need (58, p. 105).

Others are more cautious in stating their views of the relation between story-behavior and overt behavior. Murray, in his introduction to the TAT, suggests:



It may be stated, as a rough generalization, that the content of a set of TAT stories represents second level, covert . . . personality, not first level, overt or public . . . personality. There are plenty of ways of discovering the most typical trends; the TAT is one of the few methods available today for the disclosure of covert tendencies. The best understanding of the total structure of personality is obtained when the psychologist considers the characteristics of manifest behavior in conjunction with the TAT findings . . . (56, p. 16).

In similar vein, Korner states:

. . . instead of deploring the fact that fantasy and reality behavior do not necessarily correspond, as we currently seem to be doing, we can use projective techniques as a shortcut to a person's fantasy and ideational life, which then can be compared and examined in the light of his present and past actual behavior patterns (36, p. 627).

Although the relationship between covert and overt is assessed differently by various investigators, all seem to agree that the relationship is not perfect—fantasy behavior does not exactly mirror overt behavior. This omnipresent assumption can serve one of two functions depending upon the orientation of the investigator. It *can* serve simply as a convenient means of avoiding the necessity of ever being wrong. Thus, whenever inferences based on story protocols fail to relate to appropriate independent measures or observations, the clinician may simply point to the above assumption and add that only the naïve would expect always to observe linear relationships between imaginal and overt or conscious behavior. On the other hand, the investigator can use this assumption as a signpost pointing to one of the most important and difficult empirical problems facing the projective tester. This problem is the determination of the conditions under which inferences based on projective material directly relate to overt behavior and the conditions for the reverse.

It is possible to defend the position that projective techniques should not be expected to provide statements concerning overt behavior. Such a view implies that the techniques will always be used only as an "imaginal supplement" to an otherwise adequate description of the individual. Thus, given a person who "behaves" in a particular way, examination of his fantasy productions may permit us to make consistent or to account for behavior that hitherto was unaccountable. Certainly this represents an important function of these techniques. Equally certain is the fact that this is not the only circumstance under which these instruments are used. They *are* used as means of inferring overt behavior tendencies and presumably with more adequate rules of transformation they would be so used much more widely.

Investigations by Sanford *et al.* (73) and Symonds (77), which unfortunately from the point of view of sampling were both based on adolescent populations, demonstrate clearly that in some cases instead of the impulses inferred from TAT records being reflected in behavior,

their converse or opposite appear in behavior. This observation leads to the question of whether impulses that secure release in overt behavior may not need to be expressed in fantasy productions. However, the many positive relationships between imaginal impulses and overt behavior in these studies and others make it clear that it is not an either-or proposition and that the statement of the actual conditions under which the impulse is revealed or concealed must be complex.

Sanford *et al.* (73) studied the relationship in a group of school children between fantasy ratings derived from the TAT and overt behavior ratings provided by teachers who had observed the children. They found an average correlation of  $+0.11$  between the two sets of ratings indicating clearly that the fantasy ratings alone were not good predictors of overt behavior. However, there were striking differences between the different variables used in the extent to which fantasy and behavior corresponded. For some needs there was a relatively high positive relationship, while for others there was a significant negative relationship between the overt and covert. In accounting for these findings, Sanford *et al.* suggested that those tendencies which were negatively sanctioned or prohibited would be high in fantasy and low in overt behavior, while those tendencies which were encouraged by society and for which the individual could secure complete overt expression would be high in behavior but low in fantasy. High ratings would be secured in both fantasy and overt behavior for those tendencies that society encouraged but did not permit complete freedom of expression in, e.g., achievement, dominance.

Murray (56) has suggested that tendencies not inhibited by cultural sanctions are apt to be highly correlated in their fantasy and overt expression. He reports a positive correlation of over  $.40$  between fantasy and overt behavior for a group of college men on the following variables: abasement, creation, dominance, exposition, nurturance, passivity, rejection, and dejection. Negative correlations are reported between fantasy and overt for sex and no correlation between the two forms of expression for aggression and achievement. Korner (35) attempted to relate hostility as observed in a play situation with ratings of hostility in interpersonal relations with other children. She found no general relationship between the two sets of variables. Half of the children high on hostility in play situations were likewise high on hostility manifested in their dealings with other children. The remaining half who were high on hostility manifested in play situations were low on the second set of ratings. The investigator concluded that it was impossible to predict from the one situation to the other.

Symonds (77) related the fantasy themes of 40 adolescent boys and girls to adjustment ratings and teachers' ratings of behavioral characteristics. He concluded that the relationship between these two sets of variables was "insignificant and negligible."

Further evidence of the lack of a perfect relationship between

fantasy and overt behavior is provided in those cases where an individual of known characteristics fails to reveal salient aspects of himself in his TAT constructions. Tomkins (80) reports the case of an individual who had a persistent spontaneous fantasy which included as an important theme a homosexual seduction. The TAT responses of this individual gave no sign of homosexual tendencies. In accounting for this and similar cases, Tomkins suggests that the important variables are the awareness of the subject of the impulse or tendency at question and the extent to which the tendency is condoned or accepted by society. If the impulse is known and unaccepted by society, the individual will prevent its appearance in the stories he tells. If he is unaware of the tendency, it will appear in his fantasy constructions even if it is negatively sanctioned by society. Bellak (6) reports several similar instances where TAT performance fails to reveal central aspects of the individual.

Relatively little has been done in the attempt to discover and formulate signs in the stories themselves that would provide evidence concerning the probability of overt expression. There is some evidence that, as Tomkins (79) proposes, the "psychological distance" maintained by the story-teller toward the impulse or disposition in question may be an important condition relating to the degree of overt translation.

Available empirical evidence clearly indicates that the assumed imperfect correlation between fantasied and overt behavior is warranted. However, at present, we are far from an adequate formulation of the signs or cues that might permit specification from fantasy protocols alone of the behavioral tendencies that will secure overt expression as opposed to those that will not.

#### RESEARCH IMPLICATIONS

Almost all of these assumptions point to further research that would be useful in clarifying their status. Perhaps more important than research aimed at further demonstrating the warranty of these same assumptions is research that attempts to provide a more exact statement of the conditions under which the assumptions are applicable and the way in which they can be related to empirical data. For example: what are the means by which the important story in a series can be determined? How can we determine whether or not a given fantasy impulse will receive overt expression? In what way do we determine whether or not a given response has been determined by the stimulus material? What are the circumstances under which symbolic transformations must be engaged in? How do we determine the empirical referent of a given symbol? Answers to these and a host of related ques-

tions are necessary before we can hope to provide the TAT user with an explicit, repeatable set of operations for inferring motivational states.

In addition to problems connected with the interpretive assumptions and the more specific questions implied in the above paragraph, there is also the matter of formulating explicitly a method of scoring TAT protocols that is practical, intersubjective, and able to embrace a reasonable number of the behavioral variables in common use. To a large extent standardization and specific interpretive rules must wait until some agreement has been reached by most TAT users as to the major aspects of TAT response that will be focused upon in analysis.

### SUMMARY

This paper has stressed the continuity between projective testing and research and theory in other areas of psychology. In addition, it has presented ten assumptions commonly made in interpreting Thematic Apperception Test stories and has sampled briefly the empirical evidence that can be used to confirm or reject each of them.

### BIBLIOGRAPHY

1. ARNOLD, MAGDA B. A demonstration analysis of the TAT in a clinical setting. *J. abnorm. soc. Psychol.*, 1949, **44**, 97-111.
2. ATKINSON, J. W., & MCCLELLAND, D. C. The projective expression of needs: II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643-658.
3. BALKEN, EVA R., & MASSERMAN, J. H. The language of fantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neuroses. *J. Psychol.*, 1940, **10**, 75-86.
4. BARTLETT, F. C. *Remembering*. London: Cambridge Univ. Press, 1932.
5. BELLAK, L. The concept of projection: an experimental investigation and study of the concept. *Psychiatry*, 1944, **7**, 353-370.
6. BELLAK, L. Thematic apperception: failures and the defenses. *Trans. N. Y. Acad. Sci.*, 1950, **12**, 122-126.
7. BLAKE, R. R., & WILSON, G. P. Perceptual selectivity in Rorschach determinants as a function of depressive tendencies. *J. abnorm. soc. Psychol.*, 1950, **45**, 459-472.
8. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209-226.
9. BRUNER, J. S. Perceptual theory and the Rorschach Test. *J. Personality*, 1948, **17**, 157-168.
10. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. Ramsey (Eds.), *Perception an approach to personality*. New York: Ronald Press, 1951. Pp. 121-147.
11. BRUNER, J. S., & POSTMAN, L. Tension and tension release as organizing factors in perception. *J. Personality*, 1947, **15**, 300-308.
12. BRUNER, J. S., & POSTMAN, L. An ap-



- proach to social perception. In W. Dennis (Ed.), *Current trends in social psychology*. Pittsburgh, Pa.: Univ. Pittsburgh Press, 1948. Pp. 71-118.
13. CARMICHAEL, L., HOGAN, H. P., & WALTER, A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, 15, 73-86.
  14. CLARK, K. B. Some factors influencing the remembering of prose materials. *Arch. Psychol.*, N. Y., 1940, No. 253.
  15. COFER, C. N., & FOLEY, J. P. JR. Mediated generalizations and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, 49, 513-540.
  16. COX, BEVERLY, & SARGENT, HELEN. TAT responses of emotionally disturbed and emotionally stable children: clinical judgment versus normative data. *Rorschach Res. Exch.*, 1950, 14, 61-74.
  17. DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, 3, 291-308.
  18. DOUGLAS ANNA G. A tachistoscopic study of the order of emergence in the process of perception. *Psychol. Monogr.*, 1947, 61 (6), Whole No. 287.
  19. EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, 36, 34-50.
  20. ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, 64, (9), Whole No. 315.
  21. ERON, L. D., TERRY, DOROTHY, & CALLAHAN, R. The use of rating scales for emotional tone of TAT stories. *J. consult. Psychol.*, 1950, 14, 473-478.
  22. FARBER, L. H., & FISHER, C. An experimental approach to dream psychology through the use of hypnosis. *Psychoanal. Quart.*, 1943, 12, 202-216.
  23. FRANCK, KATE. Preferences for sex symbols and their personality correlates. *Genet. Psychol. Monogr.*, 1946, 33, 73-123.
  24. FRANCK, KATE, & ROSEN, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, 13, 247-256.
  25. FREUD, S. The interpretation of dreams. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud*. New York: Modern Library, 1938. Pp. 181-552.
  26. GRANDINE, LOIS, & HARLOW, H. F. Generalizations of the characteristics of a single learned stimulus by monkeys. *J. comp. physiol. Psychol.*, 1948, 41, 327-338.
  27. HAGGARD, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, 33, 257-284.
  28. HARRISON, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients: II. A quantitative validity study. *Character & Pers.*, 1940, 9, 122-133.
  29. HARRISON, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients: III. Validation by the method of "blind analysis." *Character & Pers.*, 1940, 9, 134-138.
  30. HARRISON, R., & ROTTER, J. B. A note on the reliability of the Thematic Apperception Test. *J. abnorm. soc. Psychol.*, 1945, 40, 97-99.
  31. HENRY, W. E. The Thematic Apperception Technique in the study of culture-personality relations. *Genet. Psychol. Monogr.*, 1947, 35, 3-315.
  32. HULL, C. L. *Principles of behavior: an*



- introduction to behavior theory. New York: Appleton-Century, 1943.
33. HULL, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120-134.
  34. KATZ, I. Emotional expression in failure: A new hypothesis. *J. abnorm. soc. Psychol.*, 1950, **45**, 329-349.
  35. KLEIN, D. B. The experimental production of dreams during hypnosis. *Univ. Texas Bull.*, 1930, No. 3009.
  36. KORNER, ANNELIESE F. *Some aspects of hostility in young children*. New York: Grune & Stratton, 1949.
  37. KORNER, ANNELIESE F. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619-627.
  38. KROUT, JOHANNA. Symbol elaboration test: The reliability and validity of a new projective technique. *Psychol. Monogr.*, 1950, **64** (4), Whole No. 310.
  39. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72-87.
  40. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283-293.
  41. LEVINE, J. M., & MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507-517.
  42. LEVY, D. M. Projective techniques in clinical practice. *Amer. J. Orthopsychiat.*, 1949, **19**, 140-144.
  43. LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296-309.
  44. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95-107.
  45. MAYMAN, M., & KUTNER, B. Reliability in analyzing Thematic Apperception Test stories. *J. abnorm. soc. Psychol.*, 1947, **42**, 365-368.
  46. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: an interim report. *J. Personality*, 1949, **18**, 171-179.
  47. MCCLELLAND, D. C. Measuring motivation in phantasy: the achievement motive. In H. Guetzkow (Ed.), *Groups, leadership and men: Research in human relations*. Pittsburgh: Carnegie Press, 1951.
  48. MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205-222.
  49. MCCLELLAND, D. C., BURNEY, R. C., & ROBY, T. B. The effect of anxiety on imagination. Paper read at Eastern Psychological Assn., 1950.
  50. MCCLELLAND, D. C., CLARK, R. A., ROBY, T. B., & ATKINSON, J. W. The projective expression of needs: IV. The effect of need for achievement on thematic apperception. *J. exp. Psychol.*, 1949, **39**, 242-255.
  51. MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Personality*, 1949, **18**, 236-251.
  52. MCGINNIES, E., & BOWLES, W. Personal values as determinants of perceptual fixation. *J. Personality*, 1949, **18**, 224-235.
  53. MCGRANAHAN, D. V. A critical and experimental study of repression. *J. abnorm. soc. Psychol.*, 1940, **35**, 212-225.
  54. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.

55. MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, 4, 310-329.
56. MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard Univ. Press, 1943.
57. MURRAY, H. A., & STEIN, M. I. Note on the selection of combat officers. *Psychosom. Med.*, 1943, 5, 386-391.
58. PIOTROWSKI, Z. A. A new evaluation of the Thematic Apperception Test. *Psychoanalyt. Rev.*, 1950, 37, 101-127.
59. POSTMAN, L. Toward a general theory of cognition. In J. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harpers, 1951.
60. POSTMAN, L., & MURPHY, G. The factor of attitude in associative memory. *J. exp. Psychol.*, 1943, 33, 228-238.
61. RAPAPORT, D. The clinical application of the Thematic Apperception Test. *Bull. Menninger Clin.*, 1943, 7, 106-113.
62. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, 46, 337-365.
63. RAZRAN, G. Attitudinal determinants of conditioning and of generalization of conditioning. *J. exp. Psychol.*, 1949, 39, 820-829.
64. RENAUD, H. Group differences in fantasies: head injuries, psychoneurotics, and brain diseases. *J. Psychol.*, 1946, 21, 327-346.
65. RIESS, B. F., SCHWARTZ, E. K., & COTTINGHAM, ALICE. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, 45, 700-709.
66. RODNICK, E. H., & KLEBANOFF, S. G. Projective reactions to induced frustrations as a measure of social adjustment. *Psychol. Bull.*, 1942, 39, 489. (Abstract)
67. ROSENZWEIG, S. Apperceptive norms for the Thematic Apperception Test: I. The problem of norms in projective methods. *J. Personality* 1949, 17, 475-482.
68. ROSENZWEIG, S., & FLEMING, EDITH. Apperceptive norms for the Thematic Apperception Test: II. An empirical investigation. *J. Personality*, 1949, 17, 483-503.
69. ROTTER, J. B. Thematic apperception tests: suggestions for administration and interpretation. *J. Personality*, 1946, 15, 70-92.
70. ROTTER, J. B., RAFFERTY, JANET E., & SCHACHTITZ, EVA. Validation of the Rotter Incomplete Sentences Blank for college screening. *J. consult. Psychol.*, 1949, 13, 348-355.
71. SANFORD, R. N. The effects of abstinence from food upon imaginal processes: a preliminary experiment. *J. Psychol.*, 1936, 2, 129-136.
72. SANFORD, R. N. The effects of abstinence from food upon imaginal processes: a further experiment. *J. Psychol.*, 1937, 3, 145-159.
73. SANFORD, R. N., ADKINS, MARGARET M., MILLER, R. B., et al. Physique, personality and scholarship: a co-operative study of school children. *Monogr. Soc. Res. Child Developm.*, 1943, 8, No. 1.
74. SEARS, R. R. Effects of frustration and anxiety on fantasy aggression. *Am. J. Orthopsychiat.* (In press.)
75. SIIPOLA, ELSA M. The influence of color on reactions to ink blots. *J. Personality*, 1950, 18, 358-382.
76. STEIN, M. I. Personality factors involved in the temporal development of Rorschach responses. *Rorschach Res. Exch.*, 1949, 13, 355-414.
77. SYMONDS, P. M. *Adolescent fantasy: An investigation of the picture-story method of personality study*. New York: Columbia Univ. Press, 1949.
78. THOMPSON, C. E. The Thompson modification of the Thematic Ap-

- perception Test. *Rorschach Res. Exch.*, 1949, 13, 469-478.
79. TOMKINS, S. S. *The Thematic Apperception Test*. New York: Grune & Stratton, 1947.
80. TOMKINS, S. S. The present status of the Thematic Apperception Test, *Am. J. Orthopsychiat.*, 1949, 19, 358-362.
81. WEISSKOPF, EDITH A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, 29, 379-390.
82. WHITE, R. W. Prediction of hypnotic susceptibility from a knowledge of subjects' attitudes. *J. Psychol.*, 1937, 3, 265-277.
83. WICKENS, D. D. Stimulus identity as related to response specificity and response generalization. *J. exp. Psychol.*, 1948, 38, 389-394.
84. WYATT, F. Formal aspects of the Thematic Apperception Test. *Psychol. Bull.*, 1946, 39, 491. (Abstract)

Received May 1, 1951.

## WHEN NOT TO FACTOR ANALYZE

J. P. GUILFORD

*University of Southern California*

The apparent increase in the utilization of factor-analytical methods is undoubtedly gratifying to all those who have championed those methods. Their application to research in clinical psychology as well as to research in what has traditionally been known as experimental psychology speaks well for their versatility.

It seems desirable, however, to interject some words of caution, in view of the number of misuses of factor analysis that are appearing from time to time in published articles. Many a report that includes an account of a factor analysis is faulty because of failure to take certain precautions to assure an adequate solution by that approach.

In some instances there have been poor choices of experimental variables or of populations, or of both. Sometimes it seems as if the investigator, feeling the urge to factor analyze, applies the method to his next investigation in which intercorrelations are conveniently available. Too many experimental variables have been analyzed just because they are conveniently at hand. More specifically, scores from such sources as the *Strong Vocational Interest Blank*, the *Kuder Preference Record*, Bernreuter's *Personality Inventory*, the *Minnesota Multiphasic Personality Inventory*, and the Guilford-Martin personality inventories are inappropriate experimental variables to use under most conditions of analysis. Why these variables are inappropriate will be explained in what follows.

If any research tool is to give meaningful results, it must be applied in the right place and in the proper manner. Writers on factorial methods have expressed warnings from time to time concerning requirements and limitations for the use of those methods. The most common misuses of those methods, however, cannot be entirely excused on the grounds of unfamiliarity with the literature on factor-analysis techniques. Many of the misuses and abuses could have been avoided if the investigators had observed the ordinary good rules of experimental controls and of population sampling.

The use of a complicated statistical procedure like factor analysis does not permit one to forget about the usual safeguards that should surround scientific observations. Statistical operations do not compensate for carelessness in making observations. Rather, they presuppose careful observations. They then serve as an important aid in seeing



order in the observations and in making sense of that order. Under inappropriate conditions of observation, data may appear to have an order that is misleading if not fictitious. There is no statistical magic that will give a good ordered view of nature when the data do not permit.

The discussion here will not attempt to include all the errors and pitfalls that have occurred in connection with factor analyses. In the first place, the context is limited to those factorial methods in which rotation of reference axes is an important feature. It is assumed that some rotational procedure is usually necessary to achieve an order in data that has a parallel in psychological concepts. In other words, the bias is in favor of the principle that rotations of reference axes are generally needed to yield a meaningful reference frame. It is also assumed that the order underlying a number of experimental variables is substantially simpler than the data from which it was extracted. In other words, the number of common factors is definitely less than the number of experimental variables. When the term "factor" is used hereafter, the idea of "common factor" is implied.

#### SOME COMMON FAULTS IN FACTOR ANALYSIS

1. *Too many factors are often extracted for the number of experimental variables.* Too many factors for the number of experimental variables is likely to preclude a good rotational solution. Thurstone has repeatedly emphasized that in a rotational solution the position of each reference axis must be overdetermined (8). A good rule is to have at least three variables for every factor expected. If there has been good preparatory planning of the study, one can predict the probable number of factors and whether the variables are well distributed among the factors. A paucity of variables is not only a handicap in making rotations but also in the interpretation of factors. It requires several tests all substantially loaded with a factor to indicate the generality of the factor and to define it satisfactorily.

2. *Too many experimental variables are factorially complex.* Rotations and interpretations would be much simplified if each variable were of complexity one; that is, if it measured only one common factor to any appreciable extent. This is an ideal that we achieve in test construction only once in many attempts. There is little excuse for taking almost any variable that is handy. Such variables, where there has been no effort to restrict them, are very likely to measure two or more common factors. With appropriate planning and careful test construction we can do much to reduce the complexity of tests in preparation for an analysis. The effort will pay off in terms of facilitated rotations and interpretations. It may make the difference between success or failure in achieving a solution that is acceptable.

3. *Sometimes a common factor fails to come out because it is substantially represented in only one experimental variable.* For a common factor to emerge in a particular set of experimental variables, more than one variable must have substantial variance in that factor.<sup>1</sup> It is otherwise left as specific variance so far as that particular analysis is concerned. This is ordinarily not a serious matter. It may merely mean that an opportunity has been missed to find a certain factor that is actually present.

But this is sometimes not the whole story. If this factor should be represented in minor amounts in some other variables, this common-factor variance may be extracted. But there is likely to be trouble in making rotations. If the rotating is done entirely blindly (without regard to known landmarks or psychological meaning) such a factor is frequently lost. Its variance may be divided several ways.

A recent example of this occurred in analyzing a battery of reasoning tests. A test of numerical operations had been put in the battery to assure the segregation of the number-factor variance expected in some of the reasoning tests. The number factor is so well established and a numerical-operations test is such a unique measure of it, that any solution not preserving this factor would be very suspect. Only the insistence of an axis through the numerical-operations test preserved the factor. This undoubtedly led to a clearer picture with respect to the reasoning factors.

This same fault comes up in a similar manner in connection with the analysis of scores from some personality inventories. Where some of the scores are fairly unique for single factors, those factors either do not come out in the analysis, or if weakly represented in other scores the rotational solution tends to break up their variances. This is particularly true when fewer common factors are extracted than actually are present in the scores.

4. *Not enough factors are extracted.* Experience has shown that it pays to extract a liberal number of factors. Most criteria for deciding how many factors to extract are not sufficiently liberal. No harm appears to be done in extracting too many factors. If more than the proper number of factors are extracted, the rotations will lead to the rejection of one or more of them as "residual" factors. The presence of one or two extra factors often helps to clear up the general picture of the structure for the factors that are of genuine significance.

Some investigators seem to fear that the last factors extracted are entirely a matter of error variance. It is as if they believe that up to a certain point in the extractions all variance extracted is common-factor variance and after that all variance extracted is error variance. The fact that the late-extracted variance helps to clear up the structure in

<sup>1</sup> This assumes the usual practice of using communalities in the diagonal cells of the correlation matrix.

rotations and to improve psychological meaningfulness of the factors is some evidence that true variance may be extracted late as well as early. There seems to be no good reason for rejecting the idea that error variance is also extracted early as well as late.

From this discussion it would seem to follow that the best criterion of when to cease extracting factors is the size of the larger factor loadings. As long as they are large enough to contribute something to other factors in rotation or to be built up into something psychologically meaningful they probably should have been extracted.

5. *Correlation coefficients used in analysis are often spurious.* Thurstone has often emphasized the point that the correlations used in factor analysis should be between variables that are linearly independent (8). This means that there should be no reason for covariation except that due to common factors.

There are a number of situations in psychological investigations in which specific and error variances actually contribute to intercorrelations where they should not be permitted to do so. One common situation is in connection with personality inventories in which the same items are scored with weights for more than one trait variable. This is especially true of the Bernreuter *Personality Inventory*, the *Strong Vocational Interest Blank*, and the Guilford *Inventory of Factors STDCR*. It is true to some extent of the *Guilford-Martin Inventory of Factors GAMIN*. For every item that is weighted in two scores there is a contribution to the obtained correlation between those scores. In so far as the item measures factors in common to the two scores this is legitimate. But the item's contribution to any total score also includes some specific and error variance. The specific and error variances thus contribute to the intercorrelations of items.

Let us assume that two scores were designed to measure two factors that are actually orthogonal or independent. But in the two scoring keys there are a number of items weighted similarly. There would consequently be a positive correlation, perhaps a substantial one, between scores in these two variables. A negative correlation could also be brought about between scores for two independent factors if the weights were in opposite direction when items were scored for both factors. The positive and negative intercorrelations among scores on the *Strong Vocational Interest Blank* are largely influenced by these positive and negative correlations of weights. To what extent the intercorrelations of multi-scored inventories represent actual degrees of relationship of the variables the test author intended to measure and to what extent they represent these incidental communities of specific and error variances is unknown.

The writer is often asked why some of the scores on his inventories intercorrelate so strongly when the scores were designed to measure factors. This question confuses two different things; on the one hand the factors, and on the other hand the scores which were designed to

measure those factors. Factors and their corresponding scores are logically and operationally distinct variables.

There are several reasons for the correlation of factor scores. One is the fact of intercorrelations of the factors themselves. It is very probable that some of the temperament and interest factors are actually interrelated. It is the writer's belief that at the present time we are not in a very good position to determine the extent of those intercorrelations satisfactorily. Just as there are adventitious conditions that distort the intercorrelations of experimental variables, there are also incidental disturbing conditions that give the appearance of intercorrelations of factors in an oblique solution in a factor analysis.

In addition to the genuine intercorrelations of factors themselves, there is the overlap of specific and error variances mentioned above contributing to the correlations among the scores of the factors. There is also the fact that items themselves are factorially complex. The items designed to measure factor *A*, on the whole, may be frequently involved with factor *B* and items designed to measure factor *B* may be involved with factor *A*. The factors *A* and *B* might be orthogonal but under the present level of skills in item writing it may be difficult to effect a separation in the items written to measure either factor. Even though no item were weighted for both factors in scoring, both scores would carry variance in both factors. The solution may be in the use of suppression variables (6). We should have to make some assumption about the actual factor intercorrelations, however, to assure by this method a degree of interrelation of scores consistent with interrelation of factors.

If a set of scores for factors were properly slanted so that each score measures one factor and one only, the intercorrelations of those scores would represent the intercorrelations of the factors. The factor analysis of these intercorrelations would give the second-order factors. When the intercorrelations of factor scores are distorted by all the sources of error that have just been described, however, it can be seen that an analysis will yield results that are difficult to interpret, at best.

6. *Correlations of ipsative scores are sometimes used in an analysis.* The term "ipsative measurement" was emphasized by Cattell (2). Ipsative measurements can best be defined by contrast to the more common "normative measurements." In normative measurements there is a scale for every *trait* and a population of individuals is distributed about the mean of that population. In ipsative measurements, there is a scale for every *individual* and a population of an individual's trait scores is distributed about that individual's mean. Normative scores are used to indicate inter-individual differences in a trait. Ipsative scores are used to indicate intra-individual differences in a number of traits. We compare individuals with respect to a trait by using normative scores. We compare trait quantities within an individual by using ipsative scores.



For the usual factor analysis (*R*-technique) in which the experimental variables are individual differences, normative scores are properly used. We correlate scores over a population of individuals. When the factor analysis is by the *Q*-technique, in which individuals are intercorrelated, we should use ipsative scores. The scores are then correlated over a population of traits or qualities. It is improper to use normative scores in a *Q*-technique analysis and to use ipsative scores in an *R*-technique analysis. Unfortunately, both of these mistakes are sometimes made.

One of the best examples of ipsative measurements in common practice is found in the scores from the *Kuder Preference Record*. Though these scores are not entirely ipsative, they partake strongly of the properties of ipsative measurements. The reason for this lies in the forced-choice type of item used. Where interest variables are pitted against one another rather systematically for preferential judgments, the result is that a high preference for any one interest to that extent means low preferences for other interests. It is consequently impossible for an individual to score very high in all, or nearly all, of the Kuder interest variables, as should be possible in normative scoring.

The typical profile of an individual on the Kuder inventory is distinguished by the unusual number of extreme scaled scores, high and low. This is approved by the counselor, who wants to high-light differences among the interest traits. But it represents a kind of measurement that is not suited to intercorrelations of the usual kinds where normative measurements are needed. The correlations of the Kuder scores with each other and with outside variables are of questionable meaning to the extent that they are ipsative in nature. The effect on the intercorrelations of the Kuder scores is that about two-thirds of them are negative (5, p. 615).

The scores on the Strong interest inventory have some of the ipsative quality about them, in view of the preferential items in some parts of that inventory. The effects of these parts are superimposed upon the effects of correlations of weights already mentioned.

Scores based upon the complete paired-comparison presentation of items are even more completely ipsative. The means of all individuals in the trait scores involved are equated, which would not be true in normative measurements. Such scores would be appropriate for analysis by the *Q*-technique, if the score sample were sufficiently large, but not by the ordinary *R*-technique.

7. *A pair of factors is very much confined to the same experimental variables.* This is not a very common circumstance, but one for which one should be on the alert. If factors *A* and *B* are commonly measurable by the same tests, so that in the battery being analyzed no test having variance in the one factor is free of variance in the other, factors *A* and *B* may be difficult to separate. This is especially true where an insufficient number of factors has been extracted. Sometimes one can

detect the fact that two well-known factors have thus "telescoped" into one. But how often this happens with factors that are new or unknown is hard to say. The difficulty can be forestalled to some extent by care in the selection and the construction of tests to be analyzed.

8. *The population on which the analysis is based is heterogeneous.* Although no tests of statistical significance are usually made in connection with a factor analysis, the common sampling problems have a bearing upon the success of the analysis. Too often, samples are taken from populations that happen to be most available. What is worse, different populations are thrown together without questioning what effect this may have upon the intercorrelations.

Although a number of studies have shown that the same factors may be found in the same tests when analyzed in somewhat different populations (e.g., white vs. Negro, male vs. female) with somewhat similar factor loadings (3, 6), the combining of populations is a different matter. If there are notable differences between means in both of two variables, there is a correlation between the means of those two variables. Correlations between means influence correlations of scores when such populations are thrown together.

One should not pool the data derived from different populations for the purpose of computing intercorrelations, then, unless he can show that differences in population means on the experimental variables are insignificant or unless he makes allowances for any "between-means" correlations. One procedure would be to compute the correlation matrix for each population separately. If the corresponding coefficients are similar they may be averaged to obtain a single matrix for analysis. The chief external conditions on which homogeneity should ordinarily be achieved are the familiar ones of age, sex, and education, wherever it is suspected that any of the experimental variables are related to those conditions.

9. *Not enough attention is given to requirements for correlation coefficients.* A factor analysis ordinarily begins with correlation coefficients. The results of the analysis can be no better than the data with which the analysis begins. Even elementary statistical textbooks mention the conditions needed in order that the correlation in a sample shall be a good estimate of the correlation in the population. Yet, it is apparent that many who analyze pay little attention to the question of whether those conditions are satisfied. At least, in their reports very little is said about the satisfaction of those conditions.

The least that the investigator can do is to examine the form of the frequency distribution of each variable that is correlated. The distributions should not be markedly skewed, truncated, or multi-modal. Any of these departures from unimodal, complete, and symmetrical distributions can endanger proper estimates of population correlations. The departures would probably have to be severe enough as to be obvious by inspection in order to justify concern. In other words, it would

not require statistical tests of such departures to tell whether one should do something about them. If all distributions are fairly unimodal and symmetrical, the chances are that regressions are rectilinear and that homoscedasticity would prevail in the bivariate plots.

One cannot depend upon standard psychological tests to yield the regular forms of distributions just specified for computing Pearson product-moment  $r$ 's. Distributions vary as a test is administered to different populations. The form of distribution for new and untried tests, particularly, is likely to be irregular unless pains have been taken by pre-testing to assure regular distributions.

The remedies for irregular distributions, should they occur, are common knowledge. Skewed distributions can be normalized to some standard scale, provided the raw-score scale has considerable range. They cannot be satisfactorily normalized if there are too few raw-score categories or if there is truncation. For a distribution with undue numbers of cases in end categories there is only one very convenient remedy—to dichotomize the distribution.

Dichotomizing a distribution means the computing of biserial  $r$ 's with other variables that are not dichotomized. Since biserial  $r$ 's are fairly good estimates of Pearson  $r$ 's, they may be mixed with the Pearson  $r$ 's for the analysis. Some investigators hesitate to dichotomize a distribution for computing a biserial  $r$  when the sample distribution is skewed or truncated, in view of the assumption of normality that is made about the dichotomized variable. The assumption of normality underlying the computation of a biserial  $r$  is that the *population* distribution is normally distributed. The population distribution can be normal even though the sample distribution is not, due to a faulty measuring scale. If there is no decisive information to the contrary, the population distribution on a psychological variable may often be assumed to be normal.

If more than one distribution is dichotomized, the coefficient to use is the tetrachoric  $r$  for a pair of such distributions. If the sample is fairly large (at least approximately 400) it would be convenient and defensible to dichotomize all the distributions. Whatever dichotomizing is done should divide distributions as near the medians as possible, in order to minimize the standard errors of the coefficients. Dichotomizing distributions near the medians has another virtue in connection with factor analysis. It tends to equate difficulty levels of the tests. It has been suggested that comparable difficulty levels should be achieved in order to avoid distortions in correlation coefficients (1).

10. *Difficulty levels of tests often vary substantially.* It is well known that variations in difficulty level of tests relative to the ability level of the population affect the form of distributions of scores. Inappropriate difficulty level is followed by skewing, and skewing, as was pointed out above, is followed by distortions in correlation coefficients. Certain remedies for skewing were pointed out above, but they are only correc-

tions after the fact. Scaling and dichotomizing in order to eliminate the effects of skewing appear to take care of the biases due to inappropriate difficulty levels, but they may not do so completely.

If we dichotomize a seriously skewed distribution for a very difficult test, have we effected the same kind of division of the sample as we would have done with an easier test? We do not know. We do know that a symmetrical distribution makes better use of the available range of scores. If we have a relatively short test of 20 to 25 items, for example, and if the distribution is skewed, the range of utilized score categories becomes very much restricted. With a symmetrical distribution there is opportunity for a larger range and a finer scaling.

It follows that we should attempt to avoid skewed distributions from the start. This can be achieved by good test construction and by selecting tests with optimal adjustment of difficulty level to the status of the population. If we assure regular distributions, we may then compute the more reliable Pearson  $r$ 's and avoid the recourse to scaling and dichotomizing or other remedial measures.

#### HOW THE FAULTS APPLY TO STANDARD TESTS

It was stated earlier that the scores from certain personality inventories are not suitable variables for use in factor-analysis investigations. The reasons are now much more easily presented in terms of the principles just discussed.

The scores on the *Strong Vocational Interest Blank* are not good variables for analysis because they entail the faults numbered 1, 2, 5, and 6. Covering the scope of vocational interests (also some temperament qualities that are latent in the scores) as they do, the scores may be no more numerous than the factors involved. Most of the scores are probably factorially complex, since the distinguishing qualities of a vocational group are numerous. No attempt was made to seek score categories that are functionally unique. The intercorrelations of the scores are quite generally influenced by the multi-scoring technique which contributes identical specific and error variances to two or more scores. The use of some forced-choice items introduces some ipsative properties into the scores, though this feature is probably of relatively small importance in the Strong inventory.

The *Kuder Preference Record* scores are not suitable variables for factor analysis by the usual  $R$ -technique because of faults 1, 3, and 6. Among the limited number of Kuder scores there may be more factors than scores. It is possible that one or more of the Kuder scores actually approach univocality for primary interests. If so, within the context of the Kuder score variables, these scores would yield no common factors in an analysis. If other variables were brought into the picture



those factors might emerge. The strong ipsative property of the Kuder scores, however, renders their use for intercorrelations among themselves and with other variables so questionable as to preclude attempts at analysis by the *R*-technique.

The analysis of the Guilford-Martin inventory scores by themselves in the search for first-order factors is of little use because of faults 1, 3, and 5. From previous factorial analyses of single items the indications are that there are 13 distinct factors involved. For the most part, each factor is strongly represented in only one score. For the *Inventory of Factors STDCR*, there are seriously biased scores due to multiple scoring of items. This is a minor feature of the *Inventory for Factors GAMIN*. In studies which include other variables in which any of these 13 factors are expected, it would be well to use the appropriate scores from these inventories as reference variables. Even then, it has been the policy of the writer not to include some combinations of these scores in the same analysis but to determine the correlation of a new factor with these scores after the analysis has been completed. Another procedure has been to break up the items scored for any one factor into two or three groups, each more homogeneous in appearance than the total list, and to include the part scores in the analysis. The multiple scoring of items is avoided in this procedure.

Analysis of Bernreuter's *Personality Inventory* and of the *Minnesota Multiphasic Personality Inventory*, with total scores as variables, is precluded by reason of faults 1, 2, and 5. Neither are these same scores helpful in connection with analyses of other scores because of their probable complexity and their unknown factorial compositions.

Applications of factor analysis to scores from the Rorschach instrument also encounter to a marked degree some of the difficulties mentioned here, as well as others that have not been mentioned. Many of the customary Rorschach scores are linearly dependent and distributions are usually irregular.

#### AVOIDING FAULTY STEPS IN FACTOR ANALYSIS

Ways of avoiding the faults discussed above are fairly obvious. Some of them have been mentioned in discussing the faults. It is perhaps of more value to stress good principles of "hygiene" for factor analysis than to dwell upon errors.

Requirements for effective factorial investigations have been mentioned more than once in the literature (4). Of all the general policies that might be mentioned, the one that seems to be in need of emphasizing most is that considerable careful planning should precede an analysis. It is in the designing of the particular study and later in the inter-

preting of factors that the psychologist can really show his skill as a psychological investigator. He should enjoy those aspects of his investigation most. What happens between the planning and interpretation stages can be carried out by laboratory assistants and clerks. If the genuinely psychological aspects of the investigation are not conducted with thoughtfulness and care, no amount of computational activity will make up for those deficiencies.

Briefly, a factorial study should begin with a decision as to the scope of the domain to be investigated. If one does not wish to run the risk of being involved with an unwieldy list of experimental variables, he will find it wise to select a rather limited domain for a single analysis. A domain that is not likely to involve more than 15 factors or 50 experimental variables is good policy. It is not necessary to analyze for all common factors simultaneously. The structures of the factors from limited studies can be made to fit together to complete the larger picture. There is also much to be gained from repeating analyses with variations and with combinations of domains.

The initial planning should emphasize the formation of hypotheses as to what factors are likely to be found in the selected domain and as to the probable properties of such factors. There should be no hesitation to indulge in this type of activity. Perhaps some of the more rigorously inclined investigators have avoided this approach because they have seen some of their less rigorous colleagues "dream up" hypothetical traits to accept them as established fact. Explicit hypothesis formation is something of which psychologists have done too little in their planning of research. It should make for more substantial progress for the factor analyst to begin an investigation by asking "May I assume the existence of factor *X*?" or "Does factor *Y* have these properties, or these, or these?" With properly designed experimental variables, the answer from the analysis should be of the form, "Yes, you may," or "No, you may not assume the existence of factor *X*"; "Factor *Y* is more like this than it is like that." Some investigators may be afraid of the cliché "You get out of a factor analysis exactly what you put into it." The experienced analyst only wishes that he could come closer to achieving this status of omniscience and control!

Tests especially constructed for a particular study are likely to give much clearer answers to the hypotheses than are tests already available, except where there is previous knowledge of factors in those tests. It is important for each test to be as homogeneous functionally as we can make it. It is desirable to select items for internal consistency. The usual item-analysis procedures will not assure the construction of a one-factor test. The scale-analysis methods of Guttman might

be helpful in achieving one-factor tests. The cumbersomeness of the procedure makes the cost rather prohibitive, however, and the end result may be a score that measures something entirely too specific. The best practical recourse is to the psychologist's skill in defining a hypothesized factor, in writing items, and in editing them. Ordinary item-analysis procedures will be of material help in assuring a large proportion of true variance in the scores.

In the selection of experimental variables, it is important to take into account all variances in already known factors. If one cannot exclude all common variances in known factors that he wants to leave out of the picture, it becomes necessary to put in the battery a good measure or two for each such factor. This not only identifies the factor but also segregates its variance so that it does not muddy the waters with respect to the rest of the factor structure.

With these general suggestions, most of which are not new, in mind, and with observance of the suggestions expressed or implied in connection with the various faults discussed in this paper, the one who would factor analyze will be on the road to an effective use of a very useful research tool.

## REFERENCES

1. CARROLL, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, 10, 1-19.
2. CATTELL, R. B. Psychological measurement: ipsative, normative, and interactive. *Psychol. Rev.*, 1944, 51, 292-303.
3. DUDEK, F. J. The dependence of factorial composition of aptitude tests upon population differences among pilot trainees. *Educ. psychol. Measmt.*, 1948, 8, 613-633; 1949, 9, 95-104.
4. GUILFORD, J. P. Creativity. *Amer. Psychol.*, 1950, 5, 444-454.
5. GUILFORD, J. P., & LACEY, J. I. (Eds.). *Printed classification tests*. Army Air Forces Aviation Psychology Research Program Report No. 5. Washington, D. C.: Government Printing Office, 1947.
6. GUILFORD, J. P., & MICHAEL, W. B. Approaches to univocal factor scores. *Psychometrika*, 1948, 13, 1-22.
7. MICHAEL, W. B. Factor analysis of tests and criteria: A comparative study of two AAF pilot populations. *Psychol. Monogr.*, 1949, 63, (3), Whole no. 298.
8. THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univ. of Chicago Press, 1947.

Received June 2, 1951.

## A NOTE ON CHRISTIE'S: "EXPERIMENTAL NAÏVETÉ AND EXPERIENTIAL NAÏVETÉ"

LEWIS BERNSTEIN

*Veterans Administration Hospital, Denver, Colorado*

A recently completed investigation at the University of Colorado Laboratories, the first of a series suggested by Benjamin (1), relates directly to the provocative issues raised by Christie (2) concerning the pre-experimental experience of rats. Although the results of this study are not yet ready for publication, it is felt that some of our preliminary findings should be made known at this time.

Our study was a test of Benjamin's (1) relationship-reinforcement hypothesis. More specifically, an attempt was made to establish empirically that a relationship between the experimenter and experimental animals does affect the course of learning. Special emphasis was given to the effect of an interrupted relationship upon the course of retention during a period of experimental extinction. Relationship, for purposes of this investigation, was defined in terms of the amount of handling of the experimental animals.

Fifty albino rats were used in this study. Each litter was separated according to sex, and split into three groups at the time of weaning (21 days after birth).<sup>1</sup>

Group *EH* (extra handling) was handled and petted by the experimenter for ten minutes per day per animal, from weaning through completion of the experiment. This handling was in addition to the handling required for experimental procedures.

Group *IH* (intermediate handling) was not handled until the animals were 50 days of age, at which time they were tamed by handling, for a period of 10 days. Thereafter, they were handled only as required by experimental procedures. The purpose of this handling was to reduce situational anxiety. No attempt was made to establish a relationship with these animals.

Group *NH* (no handling) was not handled at all. Special procedures were devised for introducing these animals into and removing them from the experimental apparatus without handling.

The experiment began when all animals were 60 days old. They were trained to go to the lighted side of a T-shaped discrimination box in a corrective situation, under hunger-food tension. As each animal reached the criterion of learning, it was assigned to a new subgroup, and 40 extinction trials followed. For half the animals in Group *EH*, the daily extra handling was continued throughout the extinction trials; and for the other half, handling was interrupted. Similarly, half the animals in Group *NH* continued to be unhandled while the other half received handling for the first time. Group *IH* continued to be handled for experimental purposes throughout the extinction trials.

<sup>1</sup> The relevant pre-experimental variables suggested by Christie (2, p. 333) will be reported in a future publication.



Our data for the original learning show that in terms of (a) number of trials required to master the habit, and (b) number of errors, Group *EH* was significantly superior to Groups *IH* and *NH*; and Group *IH* was significantly superior to Group *NH*.

Our extinction data show that: (a) Animals who were handled throughout both parts of the experiment made significantly fewer errors than animals who were unhandled throughout the experiment, or who received intermediate handling throughout the experiment. (b) Unhandled animals who were handled for the first time during extinction trials made fewer errors than animals who continued to be unhandled. This difference, which was not statistically significant, can possibly be explained by the fact that the handling of these animals lasted for only four days. (c) Animals who were handled throughout both parts of the experiment made significantly fewer errors than animals who were handled for the first time during extinction trials. (d) Handled animals with whom the relationship was interrupted made significantly more errors than animals with whom the relationship was continued during extinction trials. In fact, the animals with whom the relationship had been interrupted made more errors than the unhandled animals. In other words, an interrupted relationship produced more errors than a minimal relationship.

It was noted throughout the experiment that the extra-handled animals were more active and lively than the unhandled animals. This was apparent not only in maze behavior, but also in home cage activity. In the maze, the handled animals vigorously explored the apparatus—testing the wire-mesh covers, learning to open sliding doors with paws and nose, and pushing the glass food-dish aside. In marked contrast, the unhandled animals ran directly to the goal-box, without any apparent exploratory behavior. After eating the goal-food, they entered directly into the delay-chamber to await the next trial, without any exploration of the goal-box. None of the unhandled animals learned to open the doors in the apparatus. It was also noted that the unhandled animals attempted to hoard food following their daily feeding periods, whereas there was no such attempt on the part of the handled animals. On this basis, a repetition of Hunt's (4) hoarding study, introducing the variable of relationship, is being planned.

These quantitative, as well as qualitative, findings are offered as further evidence for Christie's point of view in regard to pre-experimental experience. However, these same findings suggest an alternate interpretation to Christie's (2) and Hebb's (3) emphasis on the importance of early exploratory training. Possibly a second factor was operative in Hebb's study: the pet animals may have had the benefit of a relationship with the experimenter as well as a richer experience. Of the two possible variables operative, Hebb has chosen to use the one of broader experience as an explanatory concept. It is of some interest, therefore, to note that Shurrager (5), whose animals were treated in a

fashion similar to Hebb's, emphasizes the relationship between the animals and the experimenter, rather than the concept of broader experience.

Being aware of these two variables, our animals were handled just outside their living cages in order to eliminate, as far as possible, the contamination of the relationship variable with that of broader experience. For this same reason, all animals were housed in the experimental room and none was taken outside the room. Furthermore, the position of the cages on the rack was systematically shifted twice each week, so that every cage twice occupied each position on the rack. To compensate for any additional experience which might have accrued to the handled animals by being handled outside their cages, the unhandled animals were required to run through a portable alleyway, placed between the home cages and the feeding cages, in order to obtain their daily ration. The handled animals were placed in the feeding cages by hand.

We do not mean to minimize the importance of Hebb's and Christie's emphasis on the value of early exploratory experience. We do wish to suggest, however, that in light of our findings, the following hypothesis might be tested: with breadth of early exploratory experience held constant, the group of animals with whom the experimenter establishes a relationship (through extra petting and handling) will show learning superior to that of the animals that have only broad early exploratory experience.

Finally, in connection with Christie's (2) suggestion of investigating genetic hypotheses, using animal subjects, our data show that the extra-handled and unhandled animals were equated for weight at weaning. By the age of 46 days, the mean weight gain of the extra-handled animals was significantly greater than that of the unhandled animals. Unfortunately, it cannot be stated at the present time whether this was due to better physiological use of the food consumed or to greater quantities of food consumed. We hope to resolve this question in future experiments by weighing the animals' food.

#### REFERENCES

1. BENJAMIN, J. D. Methodological considerations in the validation and elaboration of psychoanalytical personality theory. *Amer. J. Orthopsychiat.*, 1950, 20, 139-156.
2. CHRISTIE, R. Experimental naïveté and experiential naïveté. *Psychol. Bull.*, 1951, 48, 327-339.
3. HEBB, D. O. *Organization of behavior*. New York: Wiley, 1949.
4. HUNT, J. McV. The effects of infant feeding frustration upon adult hoarding. *J. abnorm. soc. Psychol.*, 1941, 36, 338-360.
5. SHURRAGER, P. S. "Spinal" cats walk. *Sci. Amer.*, 1950, 183, 20-22.

Received August 8, 1951.

## A NOTE ON THE PROBLEM OF HOMOGENEITY-HETEROGENEITY IN THE USE OF THE MATCHING METHOD IN PERSONALITY STUDIES

PAUL F. SECORD

*Emory University*

The matching method has been used in a good many personality studies and is particularly suitable for investigations utilizing data which cannot be broken down into quantitative scores or data which it is not desirable or feasible to analyze. Thus, an investigator may have judges match Rorschach test protocols with personality sketches in an attempt to determine the validity of the Rorschach test for the diagnosis of personality. One persistent problem encountered by the investigator using the matching method is that of the homogeneity or heterogeneity of the specimens being matched. The problem may be stated as follows:

In order to make the task of the judges a reasonable one, it is necessary to divide the total number of specimens into smaller groups. Each judge then matches the test specimens in each of the smaller groups with items in the corresponding criterion group; for example, he may match several sets of five handwriting specimens with corresponding sets of five personality sketches. It is obvious that lack of similarity between the specimens within any one set will make the judgments easier and greater similarity will make them more difficult. The problem, then, is one of the degree of homogeneity within the total sample of specimens and within each of the subgroups, and the relationship of this homogeneity to the conclusions which may be drawn concerning results obtained.

A study by Arnheim (1) frequently quoted with approval may be used to illustrate the fallacy of ignoring this problem. In this study, only three specimens (of handwriting) were used, those of Michelangelo, Raphael, and Leonardo da Vinci. These samples varied greatly, one being very bold, another very fine and delicate, and the other, intermediate. This heterogeneity of specimens makes matching with some other variable much easier than it would be if the specimens were similar. If the specimens representing the other variable are also heterogeneous, the task becomes still easier. An additional question here, of course, is the representativeness of the sample. With so few specimens at least one of them may correspond in character to the variable being matched with it, not because of any real relationship, but because of incidental selection of the sample. This could easily give statistically significant results when a large number of *judges* are used (2, p. 432).

The reader familiar with Brunswik's writings (2) or the note in this JOURNAL by Hammond (3) will discern that the problem under discussion is a special case of the more general problem of representativeness of the "situation" in psychological experiments. Too often situation representativeness is slighted, and only population representativeness carefully considered in psychological research.

Vernon (3) in his excellent 1936 review of the matching method, identified homogeneity-heterogeneity as the most important problem in the use of this method, but his solution to it is compressed into a single sentence and apparently has been overlooked, since all of the studies employing the matching method which have come to the attention of the writer in more recent years have either ignored the problem or have not dealt with it in a completely adequate fashion.

The solution to the problem is to treat *specimens* for matching with the same sampling precautions that *persons* in the more usual experiment are treated. The total sample of specimens must be drawn from a defined population by strictly random methods, meeting the following well-known conditions: (a) each specimen in the population must have an equal chance of being included in the sample, and (b) the drawing of one specimen must in no way affect the drawing of another. The second crucial step is involved in the formation of the subgroups of specimens. This must be done by: (a) assigning random numbers to all of the specimens by use of a table, (b) arranging the specimens in consecutive order, and (c) choosing each subgroup by taking the specimens in order of their assigned number until each subgroup is completed.

Following the above procedure will give reasonable assurance that the conclusions drawn by the experimenter can be applied to the population from which the sample was drawn, since the degree of heterogeneity in the subgroups of the sample will approach that in the population in question, provided the sample is of a reasonable size (at least 40 or preferably more). Failure to follow this procedure strictly limits conclusions to the particular sample studied, and obviously such conclusions are of little value.

#### REFERENCES

1. ARNHEIM, R. Experimentell-psychologische Untersuchungen zum Ausdrucksproblem. *Psychol. Forsch.*, 1928, 11, 1-132.
2. BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley: Univ. of Calif. Press, 1947.
3. HAMMOND, K. R. Subject and object sampling—a note. *Psychol. Bull.*, 1948, 45, 530-533.
4. SECORD, P. F. Studies of the relationship of handwriting to personality. *J. Personality*, 1949, 17, 430-448.
5. VERNON, P. E. The matching method applied to investigations of personality. *Psychol. Bull.*, 1936, 33, 149-177.

Received June 18, 1951.



# TESTS OF HYPOTHESES: ONE-SIDED VS. TWO-SIDED ALTERNATIVES<sup>1</sup>

LYLE V. JONES  
*University of Chicago*

Psychological literature abounds with experimental studies which utilize statistical tests of the significance of differences between two groups of subjects. Most of these studies present tests based upon either the distribution of Student's  $t$  or upon the distribution of  $\chi^2$ . Since the comparison of an experimental group with a control group of subjects is so fundamental to the experimental method, and since statistical tests of significance are appropriate for testing hypotheses regarding differences between two groups of subjects, it would seem important to correct a common misconception concerning the application of these tests of hypotheses.

One model for a test of significance of mean difference, the more familiar model, is that in which we test the null hypothesis,  $H_0$ , against a set of two-sided alternatives,  $H_1$ . We might formalize this test,

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0,$$

where  $\mu_1$  is the mean of the population represented by one sample and  $\mu_2$  is the mean of the population represented by a second sample. Assuming scores  $X_1$ , from the first population, and scores  $X_2$ , from the second, both to be distributed normally, and assuming the population standard deviations to be equal, we may find

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}},$$

where

$$s = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{N_2} (X_{2j} - \bar{X}_2)^2}{N_1 + N_2 - 2}}$$

and  $N_1$  and  $N_2$  are the numbers of individuals in the samples from the first and second populations.<sup>2</sup> Having stipulated a desired confidence

<sup>1</sup> This note was prepared while the writer was a National Research Council Fellow.

<sup>2</sup> Of course, if the two samples are not independently selected, we should make use of the correlation between them in the determination of  $t$ .

level,  $\alpha$ , we may enter the  $t$  table (1) with  $N_1 + N_2 - 2$  degrees of freedom and a  $p$ -value equal to  $\alpha$  to find a critical value of  $t$ ,  $t_\alpha$ . If the absolute value of the observed  $t$  exceeds  $t_\alpha$ , we reject  $H_0$  in favor of  $H_1$ ; otherwise, we accept  $H_0$ . In Figure 1 appears a distribution of  $t$  showing, graphically, the nature of this decision. This distribution corresponds to the sampling distribution of mean differences, under the null hypothesis. For any value of  $t$  to the right of  $t_\alpha$  or to the left of  $-t_\alpha$ , we reject  $H_0$ . The two shaded tails of the distribution, taken together, make up  $\alpha$  per cent of the total area under the curve.

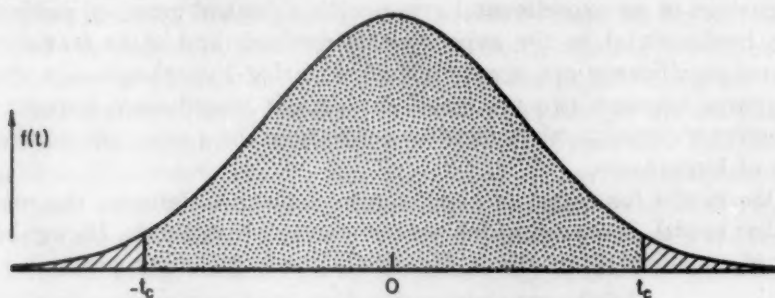


FIG. 1. THE TWO-TAILED TEST MODEL.

The model above, the test of the null hypothesis against two-sided alternatives, is the one used most often by investigators in psychology. Yet in many cases, probably in most cases, it is not the test most appropriate for their experimental problems. More often than not, in psychological research, our hypotheses have a *directional* character. We are interested in whether or not a given diet *improves* maze performance in the rat. We hypothesize that the showing of a particular motion picture to a group of individuals would lead to a *more tolerant* attitude toward certain racial minorities. We wish to test whether or not anxious subjects will respond *more actively* than normal subjects to environmental changes which might be perceived as threatening. In each case, theoretical considerations allow the postulation of the direction of experimental effects. The appropriate experimental test is one which takes this into account, a test of the null hypothesis against a one-sided alternative.

In the one-sided case we test  $H_0$  against  $H_1$ , where

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 > 0.$$

Under the identical assumptions of the two-sided model we may calculate  $t$  as before. Again a confidence level,  $\alpha$ , is stipulated. The dis-

inction between the one-sided test and the two-sided test arises in the determination of the critical value,  $t_c$ . In the present case this critical value is found by entering the  $t$  table with  $N_1 + N_2 - 2$  degrees of freedom, as before, but with a  $p$ -value equal to  $2\alpha$ . If our observed  $t$  is greater than this  $t_c$  we reject  $H_0$  in favor of  $H_1$ ; if  $t$  is less than  $t_c$  we accept  $H_0$ . The  $t$  distribution in Figure 2 exemplifies this procedure. A value of  $t$  to the right of  $t_c$  leads to the rejection of  $H_0$ , the acceptance of  $H_1$ . While the shaded area under the curve once again represents  $\alpha$  per cent of the total area, the shaded portion is restricted, in this case, to one tail of the distribution.

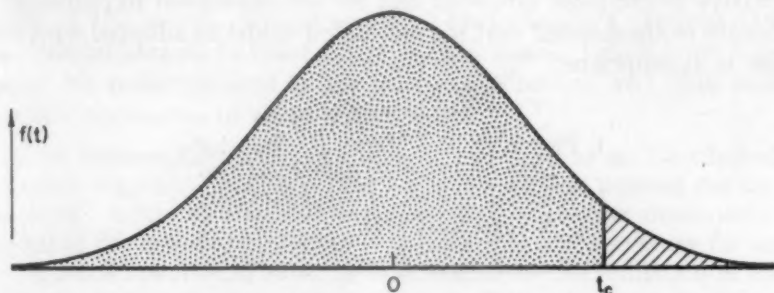


FIG. 2. THE ONE-TAILED TEST MODEL.

It might be noted that with this formulation of the one-tailed test there is no allowance for the possibility that the true difference,  $\mu_1 - \mu_2$ , is negative. In the type of problem for which the one-tailed test is suited, such a negative mean difference is no more interesting than a zero difference. In fact, the hypotheses for the one-sided case might be

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_1: \mu_1 - \mu_2 > 0.$$

In order to determine a sampling distribution under  $H_0$  we should consider the "worst" of the infinite alternatives under  $H_0$ , i.e., that alternative which would make the decision between  $H_0$  and  $H_1$  a most difficult one. Clearly, the decision would be more difficult if the true mean difference were zero than if the true difference were any negative value. Hence we would proceed exactly as in the preceding one-tailed case, utilizing, for our test, the distribution of  $t$  based upon the same sampling distribution of mean differences as before. The confidence level should be doubled to provide the  $p$ -value for entering a table to find a critical  $t_c$ , or, if it is desired to ascertain the  $p$ -value corresponding to an observed  $t$ , the correct value is one-half that given in the typical table of  $t$ .

While the one-tailed test has been exemplified here as a test of mean difference, based upon the  $t$  distribution, it is limited in application neither to mean difference problems nor to the  $t$  statistic. Indeed, wherever an alternative to the null hypothesis is stated in terms of the direction of expected results, the one-tailed test is applicable.

The failure, among psychologists, to utilize the one-tailed statistical test, where it is appropriate, very likely is due to the propagation of the two-tailed model by writers of text books in psychological statistics. It is typical, in such texts, to find little or no attention given to one-tailed tests.<sup>3</sup> Since the test of the null hypothesis against a one-sided alternative is the most powerful test for all directional hypotheses, it is strongly recommended that the one-tailed model be adopted wherever its use is appropriate.

#### REFERENCE

1. FISHER, R. A. & YATES, F. *Statistical tables for biological, medical and agricultural research*. Edinburgh: Oliver & Boyd, Ltd., 1938.

Received August 8, 1951.

<sup>3</sup> One notable exception occurs in A. L. Edwards' *Experimental Design in Psychological Research*, where the two cases are clearly differentiated.



## SOME COMMENTS ON THISTLETHWAITE'S PERCEPTION OF LATENT LEARNING

HOWARD H. KENDLER

*New York University*

Controversy frequently has an invigorating effect upon research and a confusing effect upon theory. The area of "latent learning" is an example; with each successive experiment the theoretical picture appears to become more cloudy. The time would seem to be ripe for a clarifying review of the issues. Unfortunately Thistlethwaite's attempt (14) is unsuccessful because he ignores certain basic issues. I shall avoid making a point by point rebuttal to his thirty-page review, but shall instead limit my comments to three major issues.

1. It appears flagrantly inconsistent to admit, as do Thistlethwaite and other cognitive theorists, that some reward was present during the "no food" trials of the California studies while simultaneously interpreting the results to indicate that reward is not necessary for learning. If some reward did exist, as is indicated by the reduction in errors of the Ss during the latent learning period (2, 16), as well as by the evidence of other studies (7, 10, 11), then the point at issue is not whether learning requires reward but rather how different *amounts* or *types* of reward affect learning. Admittedly the concept of reinforcement is vague at the low end of the intensity dimension. I suspect that this vagueness will disappear as the motivational component of S-R reinforcement theory continues to be developed. One must recognize, however, that it would be foolish to dispense with the concept of reinforcement because of its lack of precision at certain intensities. The concept has been too helpful in integrating large amounts of data to be discarded. Meanwhile it would be unfortunate to confuse the issue of whether learning requires reward with the issue of the relationship between learning and different amounts or types of reinforcement.

2. The underlying theme of Thistlethwaite's review of latent learning data appears to be that those studies which "demonstrate" latent learning are sound, well-designed experiments while those that fail to provide positive evidence are ill-conceived and inappropriate to cognitive theory. His analysis of the Kendler and Mencher study (9) is typical of his treatment of studies which report evidence embarrassing to cognitive theory.

1. Thistlethwaite initially criticizes the Kendler and Mencher study by suggesting that the failure to obtain latent learning could be attributed to the pronounced preference the Ss had for the black alley as compared to the white alley. No mention is made, however, of Grice's failure (5) to obtain latent learning in a similar situation with the entire T-maze painted a flat

black. Nor is any mention made of the fact that Kendler (8) reported learning of the position of food and water in a T-maze having one side painted black while the other was left unpainted. In this study, the Ss were permitted to consume the goal objects.

2. The second criticism leveled at the Kendler and Mencher study was that it used a procedure in which 50 per cent of the trials were forced. Again, no mention was made of the fact that Kendler (8) was able to obtain learning of the position of food and water in spite of his use of a 50 per cent forced trial procedure. Likewise ignored were other studies (4, 15) which demonstrated that cognitions (or habits) were established during forced trials when rewards were operating. Nor did Thistlethwaite bother to relate the results of Walker, Knotter, and De Valois (17), which he had interpreted as being favorable to cognitive theory, to his second criticism. These experimenters found no difference in the results of those Ss who had 50 per cent of their trials forced as compared with the group that was permitted to choose freely on all training trials.

3. The third and final criticism offered by Thistlethwaite was that "... the technique used by Kendler and Mencher of comparing *over the total 20 test trials* the performance of a group which found food on the same side as during the original thirst training with that of a group which found the food reversed to the opposite side, presupposed that learning is a gradual process of strengthening habits. On the basis of ... non-continuity interpretations of learning ... , one might predict that the reversed group would rather quickly give up a maladaptive hypothesis and learn within a very few trials to go to the now-correct side. The comparison of total performance scores over a long relearning period might well obscure, rather than enhance, any initial superiority of the constant group." Such a statement is misleading because it fails to mention that other response measures were reported in the Kendler and Mencher study. The criticism would not be relevant to the first test trial, the results of which fail to provide any evidence of latent learning. Nor did the analysis of the records of individual Ss provide any positive evidence. Furthermore, it seems questionable to raise such a criticism since the non-continuity interpretation of learning has been so badly discredited in recent years (3, 6). Even such cognitively oriented psychologists as Ritchie (12) and Bitterman (1) have reported data in opposition to such an interpretation. If the assumptions underlying a criticism are invalid, there is no point in offering such a criticism.

Thistlethwaite's analysis of the Kendler and Mencher study, which is typical of his treatment of findings embarrassing to cognitive theory, is neither "cricket" nor "science" for several reasons. First, of course, he fails to consider all the relevant data. It might be argued that latent learning can occur only when there is a subtle combination of several variables. The point can be a reasonable one when the necessary combination of variables is specified beforehand. Otherwise, discordant results can be attributed, as Thistlethwaite has done, to anything and everything. Second, he has ignored the fact that experiments can be designed only in terms of the *explicit* requirements of a theory. The "negative findings" of Kendler and Mencher, and others, were obtained in experimental situations which met all the *stated specifications* of cognitive theory. If the results were negative the blame belongs to

the theory. Finally, in Thistlethwaite's eagerness to perceive studies as being favorable to cognitive theory, he practically ignores the problem of why the "positive findings" of irrelevant-incentive learning have revealed such a small degree of learning. The problems, it seems to me, are for the cognitive theorist to demonstrate why such learning is so poor and for the *S-R* reinforcement theorists to show why any learning takes place. It is interesting to note the differences between Thistlethwaite's attempt to answer the first question and Spence's attempt (13) to answer the second question.

3. Finally, I would like to raise a logical point which has evaded other cognitive theorists as well as Thistlethwaite. One of the major conclusions drawn from the results of the initial California latent learning studies (2, 16) was that maze performance did not necessarily mirror maze learning. *S-R* psychologists have no argument with such a conclusion. A reasonable objection, however, might be raised because of the cognitive theorists' failure to apply this principle consistently to all aspects of their latent learning experiments. For example, Tolman and Honzik (16) have indicated that the difference between the performance of the "food" and "no food" groups during the "latent learning period" *did not reflect differences in learning*. Their reasoning was based upon the fact that both groups *performed at the same level of efficiency* following the introduction of food. If different levels of performance, however, need not imply different levels of learning, then according to the basic assumption underlying their analysis (maze performance does not necessarily mirror maze learning), it is possible that *the same level of performance need not imply the same level of learning*. That is, if one assumes, as did Tolman, an absence of an isomorphic relationship between learning and performance, it is possible that under certain conditions different degrees of learning can produce the same level of performance. The cognitive theorists have never considered this latter point but have passively accepted the equality of performance of the "food" and "no food" groups on the test trials of the latent learning experiments as evidence of equal learning.

If the cognitive theorists felt compelled to build their theoretical system around the California latent learning studies, it would have seemed advisable to test further the assumption that the similar performance of the "food" and "no food" group really indicated equal learning. An obvious experimental design would be one involving the retroactive inhibition paradigm, i.e., some activity interpolated between the training and test trials. If the results of such an experiment demonstrated that both groups resisted the effects of an interpolated activity equally well, then greater confidence could be placed in the conclusion of the cognitive theorists that the two groups had equal learning in spite of their different motivational-reward conditions. If, however, the performance of the two groups differed after the interpolated activity, then the interpretation of the cognitive theorists that

there was equal learning during the "no food" period would be suspect. Kanner,<sup>1</sup> in a recently completed experiment, similar in design to the Tolman and Honzik study, found that when reward was introduced after an interpolated activity, the performance of the "food" group was significantly superior to that of the "no food" group. Such data would be strong presumptive evidence that the Ss in the California latent learning studies, as well as in Kanner's preliminary study in which he was able to obtain latent learning, did *not* have equal learning; for if they had, their performance following the interpolated activity would not have differed.<sup>2</sup>

## REFERENCES

1. BITTERMAN, M. E., & COATE, W. B. Some new experiments of the nature of discrimination learning in the rat. *J. comp. physiol. Psychol.* 1950, **43**, 198-210.
2. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113-134.
3. EHRENFREUND, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.* 1948, **41**, 408-422.
4. GRICE, G. R. An experimental study of the gradient of reinforcement in maze learning. *J. exp. Psychol.*, 1942, **30**, 475-489.
5. GRICE, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1948, **41**, 137-143.
6. GRICE, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633-642.
7. KARN, H. K., & PORTER, J. M., JR. The effects of certain pretraining procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461-469.
8. KENDLER, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 461-469.
9. KENDLER, H. H., & MENCHER, HELEN C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, **38**, 82-88.
10. MEEHL, P. E., & MACCORQUODALE, K. A failure to find the Blodgett effect, and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, **44**, 178-183.

<sup>1</sup> KANNER, J. H. The influence of reward in the California latent learning studies. (In preparation.)

<sup>2</sup> An attempt to explain why the "food" and "no food" groups performed equally well following the introduction of food in the California studies (2, 16) as well as in Kanner's preliminary study, will be presented in Kanner's paper. It is based upon the assumption that the motivational-reward conditions of the two groups were drastically different. This assumption is, of course, quite different from the one tacitly accepted by most psychologists, namely, that both groups had the same motivation except that one group had an additional source of reward (food). It should be recognized, however, that this problem is not relevant to the present discussion. The major point of my analysis has been to question whether such "equal performance" implied "equal learning" and the results of Kanner's studies indicate that it did not.



11. REYNOLDS, B. A repetition of the Blodgett experiment on "latent learning." *J. exp. Psychol.*, 1945, 35, 504-516.
12. RITCHIE, B. F., EBELING, E., & ROTH, W. Evidence for continuity in the discrimination of vertical and horizontal patterns. *J. comp. physiol. Psychol.*, 1950, 43, 168-180.
13. SPENCE, K. W., BERGMANN, G., & LIPPITT, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, 40, 539-551.
14. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97-129.
15. THOMPSON, M. E. An experimental investigation of the gradient of reinforcement in maze learning. *J. exp. Psychol.*, 1944, 34, 390-403.
16. TOLMAN, E. C., & HONZIK, H. C. Introduction and removal of reward, and maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, 257-275.
17. WALKER, E. L., KNOTTER, Margaret C., & DEVALOIS, R. L. Drive specificity and learning: The acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, 40, 161-168.

Received June 14, 1951.

# THE BLODGETT AND HANEY TYPES OF LATENT LEARNING EXPERIMENT: REPLY TO THISTLETHWAITE

IRVING MALTZMAN

*University of California at Los Angeles*

This paper will be restricted to an analysis of the Blodgett and Haney types of latent learning experiment and to Thistlethwaite's comments upon them. Its purpose will be to show that these studies, designated types I and II by Thistlethwaite, do not demonstrate latent learning, and that their results can be interpreted within the framework of Hull and Spence's reinforcement theory of behavior.

According to Thistlethwaite, "In experiments of the Blodgett type, the question of whether there are *some* manifestations of *some* learnings during the latent period is comparatively unimportant. It is not necessary that there be no evidence of a tendency toward error reduction during the unrewarded (or slightly rewarded) trials in order for the unmanifested learning to qualify as 'latent learning.' The question of whether the abrupt increase in efficiency of running the maze upon the introduction of food is significantly greater than the comparable improvements of a control group is a statistical question. If it is in any given experiment, then that experiment may be interpreted as yielding positive evidence of latent learning" (15, p. 100).

In his review of the individual experiments Thistlethwaite does not refer again to manifestations of learning during the latent period of experiments that presumably demonstrate latent learning, viz, those of Blodgett (1), Tolman and Honzik (16), Herb (6), Haney (5), and Daub (3). In most of these experiments a large amount of improvement occurred during the latent period. As the subsequent analysis will attempt to show, such improvement has an important bearing on the interpretation of the results of these experiments.

## HAS "LATENT LEARNING" BEEN EXPERIMENTALLY DEMONSTRATED

1. *The Blodgett Design.* Blodgett's (1) comparison of the drop in errors between the control group ( $C$ ) and the three day experimental group ( $X_3$ ) gave a critical ratio of 1.338. The comparison of groups  $C$  and  $X_7$  gave a critical ratio of 1.544. Blodgett stated that the obtained differences were significant at approximately the 6 and 9 per cent levels respectively (1, p. 123). It is apparent that only one-half of the normal probability distribution function was used to arrive at these confidence levels. For both tails of the distribution critical ratios of the size ob-

tained are significant at the 12 and 18 per cent levels respectively. Blodgett did not state that he was using a single-tailed hypothesis, and he did not explicitly predict the direction of the results. The use of a single-tailed hypothesis under these circumstances may be seriously questioned. If this is the case, a significant amount of latent learning was not obtained with the criterion used.

Thistlethwaite does not present the reliability of the results, or lack of them. He does remark that Tolman and Honzik (16) obtained more significant results and that evaluation by another criterion yielded significant results (15, p. 102). He also fails to mention the results of Blodgett's experiment in terms of the time criterion, results which failed to show latent learning.

In addition to a comparison of the size of the reduction of errors or running time Blodgett employed a second criterion of latent learning. The latter method compared the drops in errors following reward of the experimental groups with interpolated drops in the control group—from *performance* levels corresponding to those from which the drops in the experimental groups began.

Blodgett reported the confidence levels as approximately the 7 and 2 per cent levels respectively. These, again, are incorrectly based on a single-tailed hypothesis. The results of this procedure are, of course, of greater statistical significance than those obtained with the first criterion used. But they do not constitute a demonstration of latent learning, at least as far as Hull's theory of behavior is concerned. The performance curves of groups  $X_3$  and  $X_7$  show some reduction of errors during the latent period. For  $X_3$  the error reduction on day three is equivalent to the effects of approximately 1.6 reinforcements in group C. Performance of  $X_7$  on the day reward was introduced was equivalent to performance in group C after approximately 2.3 reinforcements.

Hull (9) assumes that habit strength grows as a negatively accelerated exponential function of the number of reinforcements. This implies that successive increments in habit strength become progressively smaller. It follows, therefore, that the increment in habit strength after one rewarded trial in the experimental groups would be greater than the increment in habit strength after 1.6 or 2.3 reinforcements in the control group. Latent learning in terms of this second criterion can thus be readily accounted for in terms of Hull's reinforcement theory of behavior.

A third possible criterion of this type of latent learning is the saving score obtained by comparing the difference between the differences in performance of the control and experimental groups on the day reward is introduced and on some subsequent day. This scoring method would take into account initial differences in performance among the different groups, a factor that the other methods do not reflect. In terms of absolute and relative saving scores, performance of Blodgett's group C was superior to that of groups  $X_3$  and  $X_7$ . Whether or not the saving

scores are significantly different cannot be determined. However, the direction and size of such scores in all experiments of this kind indicate that they probably are significant. If latent learning had developed, and if improvement during the unrewarded period is of no importance as Thistlethwaite states (15, p. 100), there should be no difference in saving scores between the control and experimental groups.

Blodgett's experiment was repeated by Reynolds (12) who used the same six-unit T-maze pattern. In Reynolds' first study maze-wise rats were run as a control group (*C*) and as a seven-day experimental group (*X<sub>7</sub>*). Since the experimental *Ss* learned the maze prior to the introduction of reward, latent learning was not exhibited. Reynolds was criticized for using maze-wise rats (2, 15). But the fact that he also ran a control group and seven- and three-day experimental groups of naive *Ss* has been minimized by Blodgett (2), Hilgard (8), and Thistlethwaite (15).

In the second study the naive *Ss* of *X<sub>7</sub>* again learned prior to the introduction of reward producing fewer errors than the *X<sub>7</sub>* maze-wise rats. Despite this, Blodgett (2) and Thistlethwaite (15, p. 103) have argued that the maze-wise rats probably learned so rapidly because of externalization of drive.

Contrary to the impression created by Thistlethwaite and other critics of reinforcement theory (2, 8) the naive *X<sub>3</sub>* *Ss* did not learn prior to the introduction of reward. They committed 3.03 mean errors on day three compared to 2.72 mean errors of Blodgett's *X<sub>3</sub>* group on the same day. Nevertheless, these *Ss* did not show latent learning according to Blodgett's criterion. The largest drop in errors of group *C* was larger than the drop of *X<sub>3</sub>* following the introduction of reward. The same results were obtained in terms of running time.

The fact that the naive *Ss* of *X<sub>3</sub>* did not learn prior to the introduction of reward and did not show latent learning is not mentioned by Thistlethwaite. It is only indirectly implied by the criticism that repetitions of Blodgett's experiment do not get the same results because the reward used was not as palatable or the *Ss* were not as highly motivated as in Blodgett's study (15, p. 103). It is difficult to see the relevance of these criticisms, since the absolute decrease in errors or time of the experimental groups is not the criterion of latent learning. The difference in the size of the drops for the control and the experimental groups is the commonly cited criterion. Differences in palatability or strength of drive should affect the performance of control and experimental groups equally. Furthermore, it is by no means clear that all the experiments of this type, apparently showing latent learning, used a palatable mash instead of hard dog chow, or that the *Ss* were both hungry and slightly thirsty or only hungry.

Although Blodgett's experiment has been criticized as failing to



demonstrate latent learning, this by no means constitutes an adequate criticism of all latent learning experiments of this type. Thistlethwaite states that other experiments give more striking evidence of latent learning. "Blodgett used a six-unit T-maze. Tolman and Honzik used a 14-unit T-maze and obtained, following the introduction of reward, sudden drops in error and time scores which were steeper and of greater statistical significance than Blodgett's" (15, p. 102).

This statement is rather surprising because, contrary to the impression created, Tolman and Honzik did not determine the statistical significance of the drops in errors and time following the introduction of food in the experimental group as compared with the largest drop in the control group. They compared mean error and time increases of the HR and HR-NR groups, and mean drops in errors and time between HNR and HNR-R groups (16, pp. 164, 165). The first comparison indicated the effects of withdrawal of reward while the second indicated the effects of the introduction of reward as contrasted with unrewarded performance. A comparison of this sort was not made between HNR-R and HR. They did compare the performance of the two groups from the twelfth day on, following the introduction of reward on the eleventh day. In terms of errors the HNR-R group was significantly superior at the 2 per cent level. There was no significant difference in terms of time scores.

However, prior to the introduction of reward the experimental group (HNR-R) showed an amount of improvement equal to that of the control group (HR) after seven trials. If the improvement occurring prior to the introduction of food reward is of no importance as non-reinforcement theorists insist, saving scores between days 1 and 17 of the control group and days 11 and 17 of the experimental group should not differ. The control group showed greater improvement indicating that when the level of performance at the point of introduction of reward is taken into account, it cannot be concluded that latent learning is demonstrated by Tolman and Honzik.

2. *The Haney Design.* A second experimental design used to investigate latent learning was employed by Haney (5). An experimental group of hungry rats was permitted to explore a 14-unit T-maze for four days, 18 hours per day. The control group of Ss spent the same amount of time in a rectangular maze. Neither group was rewarded with food in the maze during this period. Both groups were then run hungry and rewarded one trial per day in the T-maze for 18 days. The experimental Ss that were permitted to explore the T-maze thus had an opportunity to acquire knowledge about the maze which the control Ss run in the rectangular maze could not, according to the nonreinforcement theorist.

In terms of the mean errors committed over the 18-day period the

experimental group was greatly superior to the control group. A critical ratio of 8.8 was obtained. Running times for the two groups were not significantly different.

According to the sign-gestalt interpretation the experimental Ss possessed knowledge of the sequence of paths on the first test trial, but no knowledge of the presence of food. The knowledge of "what leads to what" would therefore continue to be latent on the first trial. The experimental Ss' performance on this trial should be no different from that of the control Ss that did not have knowledge of either the sequence of paths or the presence of food. However, on the first day's run prior to commerce with the goal object the experimental group committed half as many errors as the control group (5, p. 327). Haney's experimental group showed approximately the same per cent of error reduction as Reynolds' seven-day experimental group prior to the introduction of reward. Thistlethwaite insists, and correctly so, that Reynolds' seven-day group did not meet the conditions necessary for a test of the latent learning hypothesis, because they had learned prior to the introduction of reward. The same conclusion must hold for Haney's experiment. The experimental group already learned the maze to a great extent prior to the introduction of reward.

Furthermore, the drop in errors between trials one and two for the control group was greater than for the experimental group. Latent learning was not obtained according to this criterion. Nevertheless Thistlethwaite states that latent learning was demonstrated in this experiment (15, p. 103).

Daub (3) repeated Haney's experiment in order to determine the effects that doors which prevent retracing have on performance in the test trials run in the T-maze. The criticisms of Haney's experiment apply as well to Daub's study. The control and experimental groups were not at the same level of performance at the time of introduction of reward, and a control group showed the largest decrease in errors between successive trials, although it did not occur between the first and second trial. In both studies, saving scores indicate a greater amount of improvement by the control than experimental groups following the introduction of food reward.

Another variation of the introduction of reward type of latent learning experiment has been conducted by Herb (6). A 14-unit T-maze was used in this study. Prior to training, Ss in the control and experimental groups were given six days of preliminary training totaling 30 trials in not more than two units of the maze. The control group of hungry Ss was then given one trial per day for 17 days in which food was found at the end of each blind alley. The experimental group received one trial per day for 10 days while hungry, but received no food in the maze. From days 11 to 17 food was found in each of the blind alleys entered.

The experimental group showed a large reduction in errors between the first and the tenth days in the absence of food reward. The difference was significant at the one per cent level of confidence. Following the introduction of reward in the blind alleys there was a large increase in the number of errors. Herb concluded that latent learning was evidenced by the significantly greater increase in errors committed by the experimental group on the two trials following the introduction of reward as compared with the largest increase in errors of the control group over a two-day period. It should be pointed out that a smaller difference between the two groups would be obtained if only one day following the introduction of reward in the experimental condition was compared with the largest drop in the control condition over a period of one day. Once again, here is the arbitrary selection of a criterion on the grounds that it will provide the most convincing evidence of latent learning.

The control group made a total of 145 errors or blind alley entrances on the first training day. On the first training day that the experimental group received food, day eleven, they committed 102 errors. The rate of acquisition according to Hull's theory should therefore differ in the two groups as a result of this initial difference in performance.

#### ALTERNATIVE INTERPRETATION OF RESULTS SECURED IN "LATENT LEARNING" EXPERIMENTS

In every study of the effects of the introduction of reward that was discussed, except Reynolds', latent learning was said to have been demonstrated by Thistlethwaite and the respective authors. However, it was shown that in some of these studies (1, 5) latent learning did not even occur according to its commonly accepted criterion. Saving scores indicated that none of the experiments in which this criterion could be applied clearly demonstrated latent learning. In all of these studies the performance levels of the experimental and control groups were different at the time food reward was introduced.

According to Hull's theory of behavior these different levels of performance reflect different absolute amounts of effective reaction potential (9). Since the negatively accelerated growth of reaction potential approaches an asymptote and the size of a given increment is a constant fraction of the difference between the asymptotic level and the absolute amount already acquired, the size of subsequent increments will differ for the two groups.

The different initial levels of performance, it is believed, are due primarily to the different amounts of habituation given the control and experimental groups in the training maze. In the Tolman and

Honzik experiment, for example, the experimental group received eleven more trials than the control group before the introduction of reward. Herb's experimental group had ten more trials than the control group in which to become adapted to the maze environment before food was found in the blind alleys.

This greater period of habituation permits a more complete extinction of responses incompatible with the correct responses of the training series. Estes (4) has recently shown that competing responses produce a depression in the initial phase of the performance curve reflecting the acquisition of strength of the correct response. These competing responses are probably reactions induced by fear of the new stimulus situation. Higginson (7) has shown that the arousal of fear produces an initial phase of positive acceleration in performance curves comparable to that displayed by the control Ss in latent learning experiments of the kind discussed here. Karn and Porter (10) have also indicated the importance of habituation in the determination of latent learning. Improvement in *performance* during the pre-reward period in terms of a reduction in running time or errors is not due solely to *learning* or to the acquisition of habit strength, but to the elimination of the dominant competing responses as well. The additional days of habituation received by the experimental groups afford greater opportunity for the extinction of these responses prior to the introduction of reward.

In the Tolman and Honzik study the 17 food-rewarded trials of the control group, which may also be considered habituation trials, are less effective than the 11 habituation trials and only 5 rewarded trials of the experimental group. According to the present analysis this may occur because the increments in effective reaction potential following rewarded trials are different in the two groups.

For the control group in this type of latent learning study the largest increments in habit strength which develop during the early trials must compete with the reaction potential of incompatible responses. The effective reaction potential determining the arousal of the correct responses will thereby be reduced. This would produce a positively accelerated initial phase in the performance curve. When competing responses are finally extinguished and the reaction potential determining the correct response is dominant, the size of successive increments in habit strength will be relatively small. Furthermore, competing responses will be extinguished relatively slowly because these incorrect responses will acquire some additional habit strength as a result of generalization. This is not the case with respect to the growth of effective reaction potential in the experimental group. When reward is



introduced after 11 unrewarded trials, competing emotional responses are to a great extent extinguished. As a result, the effective reaction potential determining the correct responses will not be greatly reduced by competition with the reaction potentials of incompatible responses. Therefore, the initial increments in effective reaction potential of the correct responses in the experimental group will be larger than any increment for the control group, and the rate of acquisition will be greater than that of the control group.

It is believed that the above analysis provides an account of the sudden improvement in performance following the introduction of reward in the type I and II latent learning experiments which is in accord with a reinforcement theory of behavior. It is clear, of course, that if there are adequate grounds, the account presented may be rejected or alternative hypotheses formulated without prejudicing the implications of saving scores or the reinterpretation of statistical results previously discussed.

This review has been restricted to the Blodgett and Haney type of introduction of reward experiment. The writer has attempted to show that Thistlethwaite and others have uncritically accepted the results of these experiments. The omission of any discussion of other kinds of latent learning experiment does not signify agreement with Thistlethwaite's review of these studies. Their discussion was omitted for lack of space, and because it was felt that most of them have been discussed adequately elsewhere (11, 13, 14).

## REFERENCES

1. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, 4, 113-134.
2. BLODGETT, H. C. Reynolds' repetition of Blodgett's experiment on latent learning. *J. exp. Psychol.*, 1946, 36, 184-186.
3. DAUB, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, 15, 49-58.
4. ESTES, W. K. Effects of competing reactions on the conditioning curve for bar pressing. *J. exp. Psychol.*, 1950, 40, 200-205.
5. HANEY, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, 4, 319-333.
6. HERB, F. H. Latent learning-non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, 29, 247-256.
7. HIGGINSON, G. D. The after-effects of certain emotional situations upon maze learning among white rats. *J. comp. Psychol.*, 1930, 10, 355-373.
8. HILGARD, E. R. *Theories of learning*. New York: Appleton-Century, 1948.
9. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
10. KARN, H. W., & PORTER, J. M. The effect of certain pre-training procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, 36, 461-469.
11. MALTZMAN, I. M. An interpretation of learning under an irrelevant need.



- Psychol. Rev.*, 1950, 57, 181-187.
12. REYNOLDS, B. A repetition of the Blodgett experiment on latent learning. *J. exp. Psychol.*, 1945, 35, 504-516.
  13. SPENCE, K. W., & KENDLER, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory. *J. exp. Psychol.*, 1948, 38, 106-109.
  14. SPENCE, K. W., BERGMANN, G., & LIPPITT, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, 40, 539-551.
  15. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97-129.
  16. TOLMAN, E. C., & HONZIK, E. C. Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, 257-275.

Received April 30, 1951.

## REPLY TO KENDLER AND MALTZMAN

DONALD THISTLETHWAITE

*University of Illinois*

Bertrand Russell once optimistically commented: "In science there is a technique of persuasion so effective that controversies seldom last very long. The technique consists of an appeal to evidence, not to emotions." No doubt it must seem strange to observers, both within and outside of psychology, that controversy over the Law of Effect (and more recently over its offspring, Reinforcement Doctrine) should persist so long. Part of the difficulty seems to lie in the fact that the participants in this controversy are still burdened with a number of pseudo problems. A more rigorous elimination of untestable hypotheses and continued appeal to evidence would seem to be the most promising means of bringing the controversy to an end. The critical comments of Kendler (10) and Maltzman (13) on my review (20) provide an opportunity to clarify the evidence, and possibly to adjust our differences of opinion concerning it.

### REPLY TO KENDLER

Although Kendler says his remarks are limited to "three major issues" allegedly ignored by me, he does not bother to state any of these issues. Since the basic issues Kendler has in mind are conjectural, I will limit my remarks to the specific criticisms elaborated in his paper.

1. My conclusion that some of the experiments reviewed "constitute exceptions to the alleged indispensability of reinforcement in the learning process" was derived from a consideration of the inability of reinforcement theorists to account for all of the "negative instances" accumulated (pp. 121-123). Kendler's suggestion that this conclusion is based merely upon the results of the California studies is misleading.

The charge that I was "flagrantly inconsistent" in admitting that some error-reduction occurred during the latent period of acquisition in the Blodgett kind of experiment (type I) is unjustified. It is curious that Kendler ignores my discussion of this point. I argued that before we may grant the indispensability of reinforcement for learning *we must be able to demonstrate for each instance of learning (a) that some source of reinforcement was operative in the experimental setup, and (b) that the changes in responses which are taken as evidence of the learning can be deduced on the basis of this alleged reinforcement.* Let us briefly consider the first condition. The fact that error-reduction occurred during the latent period of some of the Blodgett-type experiments indicates the operation of some motivation producing progress through the maze

during the unrewarded trials of these experiments. Kendler's conclusion from this observation is that the Blodgett type of experiment is irrelevant to the question of whether learning requires reward. His point is similar to one recently made by Spence (19). In discussing some of the irrelevant-incentive experiments (types III and IV) Spence argues that "... they have no bearing on the question of whether or not reinforcement is necessary for the hypothetical learning change to occur. All these experiments involve some form of motivation and a reinforcing state of affairs of some kind . . . Thus the question of whether or not cognitive structures can be established . . . *without a reinforcing state of affairs of any kind* cannot be answered by this experimental design" (19, p. 277). Several points need to be clarified in this argument. First of all, it is not by any means clear that *all* of the latent learning and irrelevant-incentive learning experiments involved some form of a reinforcing state of affairs. It is certainly not true that all of these experiments provided primary need-reduction contiguous with the response to be learned. Whether or not they all involved some form of secondary reinforcement is a moot question. It seems fair to ask in this regard, "How can one determine in any given experiment whether or not secondary reinforcers actually were present?" Merely to ask such a question suggests the *necessity for specifying—independently of the learning to be explained—some behavioral criteria which will be indicative of the operation of secondary reinforcement*. Until this is done reinforcement interpretations of latent learning or irrelevant-incentive learning are strictly unverifiable. If Spence and Kendler wish to insist upon "answers" to theoretical issues which are unequivocal, then it seems clear that until they make the specifications indicated they cannot expect disproof of their reinforcement assumption from *any* experimental design. The failure to confirm or unequivocally refute the hypothesis that learning requires reward does not arise from defects of experimental design but from untestable theoretical formulations. If it is true, as Hull has indicated, that "science has no use for unverifiable hypotheses" (6, p. 23) then perhaps it is time to dispense with the hypothesis that learning requires reinforcement.

I suggested in my review that we might nonetheless be interested in determining whether the hypothesis was plausible, and for this suggested especially the second of the two criteria listed above. I have explained in detail my reasons for believing that no existing reinforcement analysis satisfies this criterion. If Kendler knows of such an analysis he has not called it to our attention. It seems reasonable to conclude that reinforcement theorists are unable—despite a decade or more of concern with the problem—to formulate a consistent set of posulates from which it is possible to deduce the learning in each of these experiments. This inability itself indicates the implausibility of the hypothesis that learning requires reward.

2. Kendler's second objection was to my reasons (20, p. 115) for doubting that the Kendler and Mencher experiment disposed of Leeper's suggestion that discriminating perception of the irrelevant incentive is an important factor contributing to learning in type IV experiments. In the two immediately preceding paragraphs of my review I described the most directly relevant studies bearing upon the probable effects of asymmetry of intra-maze cues and forced trials. I did not argue that the avoidance of any *single* unfavorable condition would produce "positive results." Hence Grice's failure (4) to obtain irrelevant-incentive learning with a maze painted entirely black is not crucial to my argument. Nor is Kendler's experiment (8). On the other hand, the observations of Kendler and Mencher (11), and Meehl and MacCorquodale (14) and other unpublished results which I cited (20) clearly indicate that rats tend to develop definite preferences for black alleys and curtains over those which are white. This finding, together with that of Walker, Knotter, and DeValois (25), showing that animals with strong side preferences do not learn the position of the irrelevant incentive as well as those with "weak" preferences, strongly suggests that sharp asymmetry of intra-maze cues is unfavorable for such learning. Further, it seems somewhat disingenuous for Kendler to discount the possibility that a 50 per cent forcing technique may hinder learning when he (9) and Walker (24) have reported definite disturbances in some of their rats as a result of forcing to the nonpreferred side. If Kendler will read my entire review of type IV experiments as a unit, instead of considering parts out of context, he will find I have "related" many of the results he accuses me of neglecting.

Finally, Kendler's argument concerning the results of the continuity vs. noncontinuity experiments and their relevance to my criticism of the test procedures of Kendler and Mencher rests upon analogy. Grice's statement in connection with a similar experiment hinges upon the same reasoning: "If the animals had learned anything about the location of food, even if this expectation had not shown up on the [first] test trial . . . the Non-Reversal group would be expected to show a marked superiority over the Reversal group in rate of learning to select the path to food" (4, p. 141). My point was that any interpretation of the test series of 20 trials as providing an additional and more sensitive test of the degree of irrelevant-incentive learning has the status of a *non sequitur*. I agree that Kendler's animals gave no evidence on the first test trial of having learned the food location during the training trials. But I do not agree that comparative performances over a total of 20 test trials have any clear-cut implication regarding the existence or nonexistence of this learning. An indication of the unsuitability of the extended test treatment in this situation is provided by an unpublished experiment of mine (21). As in the Kendler and Mencher study, a reversal group and a nonreversal group were

compared in test performance following a series of training trials involving an irrelevant incentive (in this case, only the seven initial test trials given were comparable). On the first test trial 40 per cent of the reversal group as compared with 72.7 per cent of the nonreversal group selected the food path. On the second test trial 80 per cent of the reversal group selected the correct path. Furthermore, by test trial 5 the reversal group exceeded the nonreversal group in accuracy of selecting the food path. For test trials 5 through 7 the reversal group averaged 86.7 per cent food turns, as compared with 84.8 per cent for the nonreversal group. The extended test procedure over 20 trials thus promises to obscure any learning that was evident on the first test trial.

If Kendler wishes to maintain the appropriateness of including the continuity assumption as an integral part of the Kendler and Mencher experiment, then it is obvious that he cannot consistently claim that these studies were designed only in terms of the explicit requirements of Leeper's hypothesis. It seems methodologically unsound to confound an "experimental test" of one theory by introducing procedures germane to an entirely different and opposed theory.

As to the problem of why the positive irrelevant-incentive learning experiments have often revealed only a small degree of learning, the answer I suggested (20, pp. 119-120) is that the constellations of experimental conditions so far sampled have been "unfavorable." I further attempted to spell out what most likely constitutes a set of "favorable conditions." Such an attempt does not in any way entail omniscience, or ability to specify precisely the necessary and sufficient conditions for such learning. Kendler apparently believes that some of the variables listed (20, pp. 111-120) as possibly contributing to learning in type IV experiments are unimportant. A more legitimate procedure would be for him to discuss the grounds for his belief, rather than appeal merely to Occam's Razor.

3. Although Kendler's final criticism is not relevant to my review, certain obvious confusions in his argument should be clarified. It is not true that Tolman and other cognitive theorists have failed to recognize the possibility that "the same level of performance need not imply the same level of learning." Numerous experiments cited by Tolman (22, pp. 39-70) on the effects of removal of reward and change in drive indicate an awareness of this possibility. Tolman has consistently stressed that the animal's demands for the available goal objects contribute heavily to his performance. It is true Tolman and others have implied that—*motivation and reward-value of the incentive being equal*—identical levels of performance should indicate identical levels of learning. Kendler is not raising a logical point at all; there is no inconsistency involved whatsoever. Kendler is really questioning (although he explicitly mentions this only in a footnote) whether the procedure of depriving animals for the same length of time necessarily results in



equal intensities of motivation. I agree that it may not, but the burden of demonstrating the inadequacy of this commonly accepted measure of drive is Kendler's.

Finally, the logic by which Kendler argues that Kanner's results (7) constitute "evidence that the Ss in the California latent learning studies" [sic] did not have equal learning seems highly questionable. First, the generality of Kanner's results may be questioned. His experimental design was not identical with that of Tolman and Honzik—the most obvious difference is that Kanner's design provided three days of non-rewarded training as opposed to 10 days of such training in the Tolman and Honzik experiment. Schmidt (17), in a recently completed study, similar to Kanner's, found that when a rewarded test trial was given after varied delays his control groups (previously rewarded on all trials) were not superior in retention to his experimental groups (previously unrewarded on all but two of their training trials). A variable significantly affecting retention scores was number of previous acquisition trials (whether rewarded or unrewarded), suggesting that the foregoing discrepancy between the experiments of Kanner and Tolman and Honzik may be important. Finally, a point obscured in this whole argument is that it is not necessary to demonstrate "equal learning" in the California latent learning studies, but only to show that *some* learning occurred during the unrewarded trials.

#### REPLY TO MALTZMAN

Turning now to Maltzman's essay on five selected latent learning experiments, I confess that I do not know how to answer all of his criticisms. Part of my difficulty is that I cannot recognize my own statements in Maltzman's caricatures. For example, Maltzman bases much of his criticism, and presumably all of his saving score comparisons, on the argument that I claimed in my discussion of the Blodgett design that "improvement during the unrewarded period is of no importance." I deny it. The closest approximation to this that I can discover, and this removes it from the qualifying context, is: "In experiments of the Blodgett type the question of whether there is any evidence of error-reduction during the non-rewarded trials is comparatively unimportant" (20, p. 100). The subject of this sentence is "the question of whether there is any evidence of error reduction." In other words, I was talking about the existence or nonexistence of such evidence. To remove all possibility of misunderstanding, the meaning of my statement was precisely given in the immediately succeeding sentence: "It is not necessary that there be no evidence of a tendency toward error reduction during the unrewarded . . . trials in order for the unmanifested learning to qualify as 'latent learning.'" All this has evaded Maltzman, for he represents me as saying that *degree of improvement*

during the unrewarded period is of no importance. Actually, other statements made in my review (20, pp. 100, 103) clearly imply that the *degree* of such improvement is relevant and important, since the very existence of latent learning in the Blodgett design is determined by statistical comparisons of error-reductions. Further, in removing my sentence from its context Maltzman distorts its predicate. I never argued, explicitly or implicitly, that the existence of improvement (let alone the degree of improvement) was unimportant in *all* contexts. I agree that the degree of improvement during "no food" trials has an important bearing on the interpretation of the results of these experiments. I had thought this was obvious from my discussion of theoretical interpretations of these experiments. The intended predicate was of course "is relatively unimportant in ascertaining the existence of latent learning." If we put the original subject with its original predicate we get something quite different from the unrecognizable (and indefensible) statement upon which Maltzman bases much of his criticism.

At times Maltzman is talking about a definition of latent learning which is unique. In discussing Blodgett's statistical comparisons using the second of the latter's criteria of latent learning, Maltzman states: "The results of this procedure are . . . of greater statistical significance than those obtained with the first criterion used. But they do not constitute a demonstration of latent learning, at least as far as Hull's theory of behavior is concerned." This statement suggests that Maltzman is using a different definition of latent learning than was discussed in my review. For him a necessary criterion of latent learning, apparently, is that Hull's theory of behavior be unable to account for it. If Hull's theory can account for it the implication is that we do not really have a demonstration of latent learning. This is a possible definition of latent learning but it is not the one I was talking about. Maltzman has confused the statistical question of whether latent learning was demonstrated with that of whether a particular theory of learning can account for the results. The former, empirical question should be clearly differentiated from the latter, theoretical question.

With these differences in our points of departure clarified, I will consider Maltzman's claims (a) that the studies "designated as types I and II . . . do not demonstrate latent learning," and (b) "that their results can be interpreted within the framework of Hull and Spence's reinforcement theory."

1. It is possible that Maltzman does not intend to question all demonstrations of latent learning in the experiments designated as

types I and II. He has ignored the results of Wallace, Blackwell, and Jenkins (26) among the former, and those of Buxton (2) and Seward (18) in the latter. Thus, at the very best, his first contention might apply to the five experiments he discusses; it could not apply, without further demonstration, to what I have called experiments of types I and II. It is questionable, however, whether such a conclusion is tenable even with reference to this truncated list of positive latent learning experiments.

*The Blodgett Design (I).* Maltzman's criticism of Blodgett's use of the one-tailed test of significance is sheer anachronism: he accuses Blodgett of violating a convention (even yet not fully established) which had not been formulated at the time of Blodgett's experiment. It seems indisputably clear that Blodgett did entertain the hypothesis of latent learning before his results were obtained. In his review of the literature the previous experiments of Lashley, Szymanski, and Simmons are described. Blodgett did not unexpectedly discover the phenomena; he was apparently checking the results of earlier experiments, and upon confirming them invented the term "latent learning." Finally, it is not unreasonable to suppose that competent research workers could use the normal probability distribution correctly even before procedural conventions were suggested for the novice. I have already alluded to Maltzman's confusion of questions of interpretation with those of existence in his discussion of Blodgett's second criterion. It is sufficient at this point to note that his criticism (that these results do not constitute a demonstration of latent learning "as far as Hull's theory is concerned") is irrelevant.

It is difficult to determine what error scores Maltzman has used in his calculation of saving scores for experiments of type I. He says such scores were obtained "by comparing the difference between the differences in performance of the control and experimental groups on the day reward is introduced and on some subsequent day." He does not tell us how much of an interval elapsed between the former and latter performances, and whether the same interval was used for both groups. If the interval consisted of several days—and apparently it did—then such a comparison would indicate the relative efficiency of learning under rewarded trials as opposed to learning under unrewarded and rewarded trials. It would not tell us anything about the degree of latent learning demonstrated. A possible measure of the latter would be a comparison of saving scores calculated directly from error scores on the day reward was introduced for each group and on the following day. Table I shows absolute and relative saving scores, calculated by this method, for the three experiments of type I discussed by Maltzman. All of these comparisons indicate superior saving for the experimental animals, and provide further demonstration of the fact that latent learning developed during the "no food" trials.

TABLE I  
SAVING SCORES AS MEASURES OF LATENT LEARNING (TYPE I)

<i>Experimental</i>	<i>Group</i>	<i>Absolute Saving (Average Errors)</i>	<i>Saving in Per cent</i>	<i>Group Showing Greater Saving (Control vs. Experimental)</i>
Blodgett	Control	.41	14	Experimental Experimental
	X-3	1.52	56	
	X-7	1.42	62	
Tolman and Honzik	Control	1.7	18	Experimental
	X-11	2.1	38	
Herb	Control	-.2*	3†	Experimental
	X-11	-2.8*	29†	

\* Food in blinds.

† Denominators used in calculating relative saving scores are given by the expression  $14-X$ , where  $X$  equals the average number of errors on the day food was introduced.

My objection to Reynolds' use of maze-wise animals (which is distinguishable from that of Blodgett) was not because of the possibility of "externalization of drive," as Maltzman says, but because of the possibility that the response or elements of the response of running the six-unit T-maze might "have been associated with reward as a result of the previous maze experience" (20, p. 118). A more serious criticism of Reynolds' experiment—and one which applies both to the maze-wise and the naive groups—is his use of only a single motivating drive (hunger) and a single incentive (food). Maltzman says it is difficult for him to see the relevance of this latter criticism. This is an amazing confession in view of the literature demonstrating the relationships between performance and level of drive, and between performance and quality of incentive. It seems fallacious to argue that the intensity of motivation and the quality of incentive make no difference since we are only comparing the difference in the size of error decreases in the control and experimental groups. The *reductio ad absurdum* of the proposition for which Maltzman contends is that if the animals were run under a few minutes' food deprivation and were rewarded or not rewarded with a flake of food, one should get the same significant differences in error decreases as are obtained when the motivational and incentive differentials are much greater.

*The Haney Design (II).* Turning to experiments of the second type, we may briefly consider the Haney (5) and Daub (3) studies. Maltzman says, "Thistlethwaite insists . . . that Reynolds' seven-day group did not meet the conditions necessary for a test of the latent



learning hypothesis, because they had learned prior to the introduction of reward. The same conclusion must hold for Haney's experiment. The experimental group already learned the maze to a great extent prior to the introduction of reward." The last sentence is precisely correct. But Maltzman fails to recognize the point of Haney's experiment. The principal issue in these experiments was whether learning could occur under non-reward conditions. Maltzman forcefully concedes that it may. A significant difference, overlooked by Maltzman, between Reynolds' experiment and that of Haney is that in the former the previous maze experience was accompanied by reward.

Contrary to the impression created by Maltzman the commonly accepted test of latent learning in type II experiments is not Blodgett's set of criteria. The selection of the latter to evaluate these experiments might with some justification be termed an arbitrary selection made on the grounds that it will provide results least damaging to reinforcement theory. Maltzman does point to a difficulty, however, which is this: Even granting that the experimental animals in the Haney and Daub experiments learned without reward, why should not this knowledge of "what leads to what" remain latent on the first rewarded trial? Recent experiments by MacCorquodale and Meehl (12) suggest the answer. These investigators have confirmed the fact that rats after brief exploratory sessions in a maze may exhibit a tendency to avoid culs even though no incentive is available anywhere in the maze. Since we do not know the nature of the drive operating, these experiments do not unambiguously demonstrate that any drive-reduction (or reinforcement) was present during such trials. However, they suggest that the experimental animals of Haney and Daub may well have had "reason" to display their knowledge on the first test trial. The assumption made in the Blodgett design is that one can control motivation by manipulation of the incentives. However, whenever there are other important sources of motivation present, as seems indicated in the present experiments, one may question the control achieved by contrasting "food" and "no food" trials, and therefore the sensitivity of Blodgett's criteria. The statistical comparisons presented by Haney and Daub, on the other hand, unequivocally demonstrate that the experimental animals, which explored the maze, showed consistent superiority over the control animals in learning the maze. One may take this as a demonstration of a latent learning which was not manifest in the pre-test performance of these animals.

2. Maltzman has greatly simplified the task of interpretation by limiting his discussion to the results of only five of the latent learning experiments. His proposed remedy in terms of habituation processes, unfortunately, does not rest upon any clear support even in the truncated results. For example, the suggested interpretation supposes that there occurs during the "no food" trials an extinction of responses incompatible with the correct responses of the training series. But



this is clearly not consistent with the results of Herb's experiment. The responses which underwent extinction in this experiment were the finally-to-be-rewarded responses; furthermore, responses incompatible with these were *increasing* in frequency during the "no food" trials. Finally, it is not clear how the postulated differential extinction (habituation) occurs. It might possibly occur through differential reinforcement. However, no source of reinforcement during the "no food" trials, or during the exploratory periods, has been identified, nor has any alleged source of reinforcement been shown to be actually present and active in the unrewarded trials of these experiments (1, 2, 3, 5, 18, 23, 26). It is therefore gratuitous to argue that these results have been shown to be interpretable within the framework of Hull and Spence's reinforcement theory of behavior.

## REFERENCES

1. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, 4, 113-134.
2. BUXTON, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theor.*, 1940, 2, No. 2.
3. DAUB, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, 15, 49-58.
4. GRICE, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1948, 41, 137-143.
5. HANEY, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, 4, 319-333.
6. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
7. KANNER, J. H. Do non-rewarded animals learn as much as rewarded animals in the California latent learning studies? *Amer. Psychologist*, 1951, 6, 266. (Abstract.)
8. KENDLER, H. H. The influence of simultaneous hunger and thirst upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, 36, 212-220.
9. KENDLER, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, 40, 265-270.
10. KENDLER, H. H. Some comments on Thistlethwaite's perception of latent learning. *Psychol. Bull.*, 1952, 49, 47-51.
11. KENDLER, H. H., & MENCHER, HELEN C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, 38, 82-88.
12. MACCORQUODALE, K., & MEEHL, P. E. On the elimination of cul entries without obvious reinforcement. *J. comp. physiol. Psychol.*, 1951, 44, 367-371.
13. MALTZMAN, I. The Blodgett and Haney types of latent learning experiment: Reply to Thistlethwaite. *Psychol. Bull.*, 1952, 49, 52-60.
14. MEEHL, P. E., & MACCORQUODALE, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, 41, 372-396.
15. MEEHL, P. E., & MACCORQUODALE, K. A failure to find the Blodgett effect, and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, 44, 178-183.
16. REYNOLDS, B. A repetition of the Blodgett experiment on "latent

- learning." *J. exp. Psychol.*, 1945, 35, 504-516.
17. SCHMIDT, HANS, JR. The acquisition and retention of latent learning. Unpublished M. A. thesis, Univ. of North Carolina, 1951.
  18. SEWARD, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, 39, 177-186.
  19. SPENCE, K. W. Theoretical interpretations of learning. In Stone, C. P. (Ed.), *Comparative psychology* (3rd ed.). New York: Prentice-Hall, 1951.
  20. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97-129.
  21. THISTLETHWAITE, D. Drive discrimination and food location as factors in irrelevant-incentive learning. *Amer. Psychologist*, 1951, 6, 287-288. (Abstract.)
  22. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century, 1932.
  23. TOLMAN, E. C., & HONZIK, H. C. Introduction and removal of reward, and maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, 257-275.
  24. WALKER, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, 38, 39-49.
  25. WALKER, E. L., Knotter, Margaret C., & DEVALOIS, R. L. Drive specificity and learning: the acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, 40, 161-168.
  26. WALLACE, S. R., JR., BLACKWELL, M. G., JR., & JENKINS, I. Prereward and postreward performance in the "latent learning" of an elevated maze. *Psychol. Bull.*, 1941, 38, 694. (Abstract.)

Received September 15, 1951.

### CORRECTION

A. RABIN AND W. GUERTIN

*Michigan State College*

In our paper entitled "Research with the Wechsler-Bellevue Test" in the May, 1951 issue of *Psychological Bulletin* reference was made to Dr. Kutash's research report (Reference No. 85). At the time that this study appeared it was of considerable importance in showing the necessity for caution in interchanging the Stanford-Binet and Wechsler-Bellevue tests. Through this paper as well as others the importance of the chronological age in connection with allowances for deterioration after maturity has become recognized.

At the time of the writing of this review our emphasis was on the comparability of the Stanford-Binet and Wechsler-Bellevue with respect to individuals who are in the presenescent period of life. This led us to refer to his article with such phrases as "did not control age" and "uncontrolled age factor." This presented Dr. Kutash's work in an unfavorable light and we hereby recommend that this be kept in mind in reading this review. Actually, Dr. Kutash's work showed a careful control of all relevant variables which were under consideration including that of age.

## BOOK REVIEWS

BLAKE, R. R., & RAMSEY, G. V. (Eds.) *Perception: An approach to personality*. New York: Ronald Press, 1951. Pp. viii+442. \$6.00.

This book consists of a series of lectures more or less as delivered by eleven well-known psychologists and the late Alfred Korzybski at the 1949-50 Clinical Psychology Symposium held at the University of Texas. The first impression it creates is one of brilliant confusion, as if one were catching fascinating glimpses of a problem, many of which are in themselves impressive but all of which together do not create any kind of an over-all picture. In fact, I am tempted to complain that we have here one further example of the disturbing tendency in American psychology to substitute a *collection* of interesting ideas for hard theoretical thinking. The editors attempt to remedy this lack of synthesis somewhat by an introductory chapter which, however, does little more than add some new terminology for the already overburdened reader and provide a brief abstract of what each lecturer said.

It is difficult for the reviewer to avoid doing much the same thing, since it is nearly impossible, as in all symposia of this sort, to make generalizations that apply to all the authors. Yet, on the other hand, it is equally impracticable to review each of the twelve lectures individually. As a kind of compromise I will attempt a few generalizations anyway and, in the course of discussing them, try to comment on the contribution of each of the authors.

My first generalization is that the lecturers are, by and large, excellent representatives of what might be called "progressive" or at least "new" trends in American psychology. In fact, I have just finished warmly recommending the book to a group of graduate students in Europe as a means of getting an idea of what is happening in American psychology. One might suppose that the subject matter of the symposium is too narrow to warrant such a recommendation but as a matter of fact the topics covered range over the whole field of psychology. Thus we have Morgan on the nervous system, Beach on hormones, Rogers on Q-technique and nondirective therapy, Bronfenbrenner on a comparison of personality theories, Korzybski on semantics, Frenkel-Brunswik on case studies (presented in almost tedious detail) of the perceptions of adolescents, Hilgard on learning in perception, etc. In fact, many of the lectures have little to do with perception as such, but if one gives up the idea that they should and reads each one for itself as typical of some new and interesting development in psychology, one soon ceases to worry about this. A serious complaint is sure to be raised, however, about an omission in what otherwise seems so complete a coverage of the subject of perception. There is almost nothing in this

book about the "stimulus"—nothing much in it, for example, except incidentally in Hilgard's chapter on learning in perception, about the classical treatments of perception as represented, say, in Gibson's or Stevens' work. This may appear strange until it is realized that the symposium was clearly designed to stress the internal rather than the external determinants of perception. In my opinion, such an emphasis should still not have excluded altogether a lecture on external determinants (perhaps in their relation to internal determinants), particularly in view of the inclusion of a lecture on the influence of the nervous system on perception. Gestalt psychology is mentioned on only one page according to the index, a fact which highlights this error of omission as well as any.

To sum up our generalizations so far: the book is frequently brilliant but as a whole unintegrated, representative of many new trends in American psychology, extremely broad in its coverage but stressing the personal rather than the external or situational determinants of perception. What else can be said about it as a whole? One other immediate impression is that much of it is admirably written and easy to read—possibly because the material was prepared originally to be spoken. Morgan's chapter on "Structural factors in perception," Beach's chapter on "Body chemistry and perception," Bruner's chapter on "Personality dynamics and the process of perceiving," to mention the first three that come to mind, are models of expository clarity. On the other hand, the book also contains some examples of that love of words and connotative distinctions that has so plagued personality theory in chapters by Cameron, Korzybski, Klein, and others. This difference in styles of writing seems also to reflect a deeper difference in modes of thought. One can see evidences still of the traditional cleavage between the "tough-minded" experimentalists and the "tender-minded" clinically-oriented psychologists.

Particularly interesting in this connection is Bronfenbrenner's spirited defense of "clinical" thinking. He challenges sharply "the dubious assumption that scientific wisdom increases by steps significant at the 5 per cent level" and insists on the importance of "vague gropings" for which, he argues, we American psychologists have had to rely chiefly "on our European-trained colleagues." He then goes on to condense and compare in propositional form the personality theories of Lewin, Freud, Rank, McDougall, and Sullivan—surely a breathtaking task! I am greatly impressed by the job he has done and feel sure I will be sending students to read this chapter for years to come, but two kinds of questions kept recurring to me as I read through it. First, why is it that Beach in writing on hormones doesn't also find it convenient to organize his presentation around the contributions of one man after another? Is it a sign of maturity in a science when our thinking is organized around problems rather than around "great men"?

Why, after all, should we bother with all the ideas any one man has, when many of them are undoubtedly bad? Secondly, why is it that Bronfenbrenner finds a brief clinical analysis of the man so useful in understanding his theory? Thus he discovers the genesis of Rank's ideas in his relationship to Freud, of Freud's ideas in his relationship to his father, etc. This gives one the queer, and rather shocking, impression that a personality theory is a kind of gigantic projective system. And in fact this seems to be the case so long at least as we do not have data—"autochthonous" empirical data—which minimize the relative importance of the internal, personal determinants of such scientific belief systems.

Therefore, while I find Bronfenbrenner's analysis historically fascinating, the most important and interesting parts of this book for me are those written by men about their own research data. The capacity of men to invent new "symbol systems" or connotative theories has always seemed to me so infinite that new ones leave me a little cold (as does Cameron's in the present book), unless they are closely tied to actual experimental data. Nowhere is the contrast I am trying to make clearer than in Klein's chapter which reflects remarkably divided loyalties. In the first part of it, he revolts against all sorts of traditional psychological concepts, such as the nomothetic approach in general or such commonplace terms as "traits" and "habits" in particular. Thus he wants to replace a term like "perceptual trait" with "matter of course avenues" or "perceptual Anschauungen." He rejects the classical distinction of introversion-extroversion (or even extratensive-introversion) in favor of "self-inward" and "self-outward." And his discussion is so loaded with new words and distinctions that one might at this point classify him with the "connotative" theorists who seem to be interested in spinning out differences that don't make an operational difference. On the other hand his actual research discussed later in the chapter is imaginatively conceived, carefully executed, and intensely interesting. My own reaction to it is that at last someone has begun to isolate distinctive modes of perceiving the world which different individuals characteristically adopt in attempting to adjust to it. Yet somehow all the new terminology Klein introduces to explain such fascinating results seems largely unnecessary. Why give up a perfectly good commonplace term like "perceptual trait" just because someone used the term in another sense once? Thus Klein gives the impression of having one foot in the "clinical camp," where personality concepts are still rather like "projective" products, and another foot in the experimentalist camp where they are based more firmly on external, empirical facts. Perhaps only in Bruner's stimulating chapter is there evidence of a genuine synthesis of the contributions of both viewpoints into a tentative theory which avoids the "blindness" of the empiricist on the one hand and the "projections" of the clinician on the other.



Another general impression of the book is that many of the chapters represent useful condensations by the authors of longer books. Thus, Morgan summarizes some of the most important ideas in Morgan and Stellar; Beach abstracts from his recent book written in collaboration with Ford; Korzybski writes entertainingly about ideas that he has treated extensively elsewhere and manages in the process to display many of the bad language habits he so passionately objects to. Rogers, Cameron, and J. G. Miller likewise contribute condensations of longer works. In a somewhat different class are the condensations of *future* works which we will await with heightened interest from Bruner and from Frenkel-Brunswik. Finally, there are useful summaries of research conducted by others—such as Hilgard's discussion of the Ames' demonstrations and Dennis's extremely valuable summary of Piaget's research on children's ideas. Thus one can readily imagine using the book as the basis for a seminar or as the starting point for more detailed discussion of original works.

Finally, the book has a good index which, if properly used by students or in seminars, should aid in producing discussion and even integration of the differing viewpoints presented. Take the Rorschach, for example. If you look up the references under this heading in the index, you will find Dennis castigating psychologists quite properly for accepting without question the fact that a color response to an inkblot means something about emotion when the evidence supporting the association is scarcely better than that for connecting red hair with emotion, Klein going behind the Rorschach to the "perceptual attitudes" which produce responses to it, and Elsa Frenkel-Brunswik discussing the place of Rorschach in the historical development of the study of personality. The index will also show up some interesting contradictions, as, for example, under "Frequency" where there is a reference to Bruner asserting that an hypothesis will be stronger the more frequently it has been confirmed and another to Hilgard arguing that this is not always true. The student can learn much in the course of resolving such a disagreement (which in this instance is more apparent than real). In short, the book contains many new and interesting ideas presented by some of the best psychologists in the business today and should provide an excellent basis in its very variety and inconsistency for serious discussion and theorizing for a number of years to come.

DAVID C. McCLELLAND.

*Wesleyan University.*

FORD, C. S., & Beach, F. A. *Patterns of sexual behavior*. New York: Harper, 1951. Pp. viii+307. \$4.50.

Here is sex in cultural and evolutionary perspective. In writing this book, Ford and Beach have accomplished a difficult task. In one

brief volume, they have summarized the major facts of mammalian sexual behavior, have laid the groundwork for future scientific studies and theories of sexual behavior, and have spelled out what the layman ought to know about sex. No student of behavior, comparative or physiological, social or abnormal, ought to overlook this work. No person, married or about to get married, should miss reading it.

Do not let the clear and simple writing of these authors make you believe that their book is lightweight or shallow. Although it is not peppered with references, it is a well-documented volume. It is written with the utmost scientific caution. Yet it offers interpretation. Best of all, it has a point which the scientist as well as the layman needs to learn: how to be objective about sexual behavior without depending only on cold, mass statistics.

The main scientific problem that Ford and Beach face is to determine which aspects of man's sexual behavior are the product of experience and training and which are determined by his physiological make-up. To accomplish this end, the authors evaluate sexual behavior in American society from the cross-cultural and comparative or evolutionary points of view. Practices and customs in 190 cultures are reviewed. Studies of more than thirty species of mammals from the shrew through the chimpanzee are presented. In addition are included direct experimental investigations of the roles of physiological mechanisms and the learning process in sexual behavior.

These four approaches in the analysis of sexual behavior provide excellent criteria for the evaluation and understanding of sexual practices and attitudes in American culture. In simplest terms, they amount to this: (a) Any practice that is widespread among the different cultures and also appears in lower mammals, particularly the subhuman primates, can be suspected of being a basic physiological pattern. (b) Any behavior that is practiced in some cultures and not in others and that does not appear among the lower mammals is probably the outcome of learning. When they apply these criteria, Ford and Beach follow a simple formula.

1. They take up the occurrence of the type of sexual behavior in American society and the prevalent attitude toward it. A good example is their treatment of homosexual behavior.

2. They examine other cultures which have different attitudes and practices in regard to the behavior in question. In the case of homosexuality, they report that even the most restrictive cultures have some incidence. Furthermore, some cultures are permissive and even demand that certain individuals practice homosexuality at least during some period in their lives.

3. They review the evidence on subhuman primates and subprimate mammals. Here they find that homosexual behavior does occur, that mammals are basically bisexual, and that which pattern comes out depends upon the nature of the immediate stimulation and, possibly, the past experience of the individual.

4. They examine the direct physiological evidence that the behavior is determined by the genetic makeup of the individual. In the case of homosexuality, they find that male and female hormones do not induce masculine and feminine behavior, but rather simply excite the organism to either pattern depending upon other circumstances.

5. They review studies designed to get at the role of learning. As far as homosexuality is concerned, they can show that rearing rats in isolation or segregation does not alter the incidence of homosexuality and heterosexuality compared to what is found in rats reared in cohabitation. Some evidence indicates that monkeys that are social outcasts may develop homosexuality, but it is also true that dominant males that have ready access to females may practice homosexuality in addition to heterosexual copulation.

6. Comparisons of males and females of all species and cultures are made. In this case, it is found that overt homosexual acts are much more prevalent in all of these cases in the male than they are in the female.

7. The developmental approach is also utilized. Here it is shown that homosexuality occurs much more universally in immature organisms among both animals and men than it does in the sexually mature members of the species.

8. Finally in their analysis, the authors offer an interpretation of the behavior under consideration. For homosexual behavior, they say that it is a natural type of sexual expression, dependent fundamentally on the fact that mammals, including man, have the kind of nervous systems which permit them to be aroused by masculine and feminine types of stimulation and to execute many of the components of both masculine and feminine mating patterns. It is clear that hormonal abnormalities are not typically at the basis of homosexuality, but that either hormone can serve to excite both patterns. Homosexuality occurs most commonly in sexually immature organisms, but the typical pattern for mature individuals among all animals and all cultures is heterosexual copulation. The decrease in homosexuality in the mature animal is probably largely the result of the competition offered by the heterosexual mode. Among humans, there is also the fact that many societies discourage or violently oppose homosexuality in the early training of the individual. Because of man's tremendous capacity for learning and symbolism, his homosexuality may differ from that of the lower forms. For example, completed anal copulation or the use of penis substitutes are probably impossible among lower mammals and occur only rarely in the great apes. It is also true that only among humans is there any noteworthy incidence of cases where homosexuality is more or less permanently preferred to heterosexual behavior.

The authors make the same kind of complete analysis of other aspects of sexual behavior: ways of attracting mates, the nature of foreplay and types of heterosexual stimulation, the positions and techniques of coitus, the physical and psychological circumstances under which copulation will occur, the occurrence of single and multiple sexual partnerships, masturbation, and relations with other species. Each topic is given a whole chapter, and in addition, chapters are devoted to the development of sexual behavior in the individual, the relation between sexual behavior and feminine fertility cycles, and the relative

importance of hormones and central nervous mechanisms in the control of sexual behavior. An introductory chapter discusses the nature and limitations of the task undertaken in the book, and the final chapter is a good summary, viewing human sexual behavior in perspective. A good glossary is included at the end of the book.

A number of important generalizations are made in the final chapter as well as elsewhere throughout the book.

1. Changes have taken place in evolution which have released the primates, especially man, from strict hormonal control of sexual behavior.

2. The higher an animal on the evolutionary scale, the more its sexual behavior is dependent upon the cerebral cortex and the more it is influenced by learning.

3. Certain aspects of man's sexual behavior are clearly the results of his mammalian heritage: sexual play in childhood, masturbation, homosexuality, and many aspects of sexual foreplay, including mutual manipulation of the genitals with the hands and mouth and general grooming behavior.

4. Some of the similarities seen in different cultures seem to be the result of common learning experience rather than any biological mechanism common to all men. For example, while incest taboo is almost universal in human cultures, there is no good evidence for it among animals, and there are few sound biological indications against it. The authors suggest that incest regulations are designed to preserve the nuclear family group in societies by keeping it free from the dangers of sexual jealousies and conflicts.

5. It is amply clear that many of the sexual practices of man, found only in particular cultures or even only in particular individuals, are the product of social learning, very often a deliberate part of the training of the human individual.

Taken in perspective, this book is a tremendous achievement in the fields of comparative psychology and social anthropology. It is an excellent example of what these approaches can offer in other areas of human behavior as well as in the area of sexual behavior. But as the authors point out, it is only a start, really a primer. Much remains to be done, but the groundwork has been laid.

ELIOT STELLAR.

*The Johns Hopkins University.*

MORGAN, CLIFFORD T., & STELLAR, ELIOT. *Physiological psychology* (2nd Ed.). New York: McGraw-Hill, 1950. Pp. ix+609. \$5.00.

The authors of this edition have produced a distinguished successor to the original. Its coverage is broad, and for a technical book its reading style simple. In general, accuracy has not been sacrificed for readability. This edition continues to be the best available single treatment of the subject for teaching purposes. It will undoubtedly influence another generation of students exposed to this field of study.

The text has been considerably revised. Almost every area previously covered has been freshened by an evaluation of more recent experi-

mental findings. Theoretical speculations (including those developed previously by the senior author) have been sharply reduced. A good deal of the psychophysical material has been omitted. The treatment of motor functions, color vision, and learning has been extended.

It would be easy to disagree with the authors on their choice and organization of the material. That there is room for disagreement is attributable to the breadth of the field, forcing selection on those who are courageous enough to undertake the task of presenting an objective and balanced overview of physiological psychology. The authors have been highly (but not uniformly) successful in accomplishing this task. A weakness arises out of the contradiction between their avowed main interest as the study of man (p. 502) and a heavy emphasis on the results of experiments on lower animals (in the interest of objectivity). The growth of knowledge about the main object of interest, the human organism, is thereby deferred by a preoccupation with organisms on which psychological data as intervening variables are extremely difficult to study. As a consequence, the "givens" are usually some structural insult followed by a *limited* description of affected behavior. The "in-betweens" are neglected. The temptation is strong (and frequently not resisted) to fill the void with overaged reductionist assumptions.

The reader is often led to believe that behavior modification resulting from brain extirpation means that the part extirpated is responsible for the particular behavior. If this logic were applied to the optic chiasm we should consider it responsible for vision! Obviously many other structures, and experiences as well, participate in making visually dependent behavior and experience possible. Because the authors limit the description of behavior when discussing correlated physiological events, the reader is permitted to infer that the control or responsibility for direction of an activity belongs to these events. Note, for example, the uncritical use of the words "control," "executive function," "responsible," among others, in the following quotations:

The central nuclei of these nerves (III, IV, and VIth cranial) make up an important center for the control of eye movements. . . . (p. 24).

Primary control of autonomic functions is vested in the hypothalamus (p. 43).

This (the corticobulbar tract) serves the motor nuclei . . . of the cranial nerves that control movements of the eyes, face, lips, tongue and vocal cords (p. 283).

. . . it is fair to conclude that area 4 is the principal executive center for motor coordination (p. 329).

Do the authors mean, for example, that the motor nuclei of the cranial nerves *direct* the movement of the eyes, face, lips, etc.? Or do they mean that these nuclei are among the structures and experiences which make all of the activities of these organs possible?

Consider also the statement: "We must ultimately understand



nest building in terms of thermal changes on the nervous system" (p. 414). Apparently the authors do not mean by "ultimate understanding" a complete description of *all* of the processes involved in this activity but rather they imply a hope that some day it will be possible to demonstrate that thermal changes in the nervous system trigger nest building.

Similarly, summarizing the results of Krechevsky's work on brain mechanisms and hypotheses in rats they state: "... a cortical area corresponding fairly exactly with the striate (visual) area was responsible for the visual hypothesis" (p. 492). Is the production of visual hypotheses in rats a unitary psychological function emanating from the striate area or is the presence of the striate area critical for this function?

In line with the reductionist bias of the authors, the study of spinal or other reduced preparations is justified (p. 299) without cautioning the reader that the relevance of these studies to an understanding of the behavior of the intact organism is questionable. For example, the crossed extensor of the spinal cat and the reflex figure of the decerebrate cat are alleged to be basic to walking and running (pp. 301-304). The claim is made in spite of the fact that neither the conditions for eliciting them nor the actual form of the movements themselves is comparable to those for walking and running. It is highly doubtful that these reflexes have anything to do with walking and running.

Apart from disagreement with the authors on these theoretical issues, the reviewer believes this edition is superbly well done.

JOSEPH E. BARMACK.

*The City College of New York.*

BARTLEY, S. HOWARD. *Beginning experimental psychology*. New York: McGraw-Hill, 1950. Pp. viii+483. \$4.00.

This is another of the recent books aimed at an undergraduate course in experimental psychology. In particular this is a beginning book about experimental psychology for use in a lecture course without accompanying laboratory. As stated by the author, "Although laboratory work is indispensable at an early stage of your training, your first introduction to experimental psychology is well accomplished through verbal description and explanation provided by textbook and lectures."

It may be questioned whether this aim can be achieved without practical work in the laboratory. To be sure there is much debate concerning the most effective manner of teaching *experimental method* as applied to psychology. Too often the traditional course has become routinized so that students follow directions blindly with little appreciation of the nature of experimentation, the design of experiment, the application of controls, and the very essence of scientific observation. Without a laboratory course in which the student can apply these

principles, however, a course in experimental psychology can become merely another subject-matter course about experimental psychology.

The author gives four characteristics of his book. Each of these will be discussed in turn.

1. "*Short enough to be covered in one quarter or one semester.*" There are 48 short chapters, each of single assignment length. In most chapters one or two typical experiments are presented in detail although some chapters follow the typical textbook form. Review questions for the student are found at the end of each chapter. Although the book is designed for a semester course, some doubling up of assignments would be necessary to fit forty-eight sessions into a typical semester with three meetings per week.

2. "*Dealing with principles by way of illustration rather than being a compendium of the findings of experimentation.*" Because of the arrangement into many short chapters the student is presented with a kaleidoscope of experimentation and methodological detail. For the beginning student the significance of these details may be lost unless he can be made to see the relevance, not only of the details of the experiment but also of the particular experiment, to psychological knowledge and theory in general. In some of these chapters there is a dearth of such background material. On the other hand, some, like Chapter 41 on "The Social Basis of Perception," are well done in this respect.

Certain of the experiments are rather complex in design and indecisive in outcome. The author does not always make clear the steps by which certain general theoretical statements are derived from experimental particulars. For example, the following is one of the conclusions of the chapter on configurational learning in the goldfish. "Perkins also saw in the behavior of the fish the exemplification of the 'law of least action.'" This is the first mention of the law anywhere in the text. There is no further statement or elucidation of the law except for a footnote as follows: "You as a student of experimental psychology, will do well to look into the concepts of least action, maximum work, etc., formulated by Wheeler. They are given in his textbook, *The Science of Psychology*."

3. "*Representing all the major areas in psychology rather than just one or two which happen to be the central interest of the author.*" By this the author undoubtedly refers to the somewhat broader selection of fields than is usual in the traditional texts on experimental psychology. In addition to sections on History, The Nature of Experimentation, Psychophysical Methods, The Broad Features of Sensory Experience, Physiological Psychology, Comparative Psychology, Learning and Memory, and Preparedness and Adequacy, there appear five sections covering, respectively, Social Psychology, Legal Psychology, Child Psychology, Clinical Psychology, and Industrial Psychology.

4. "*Designed so that the assignments may follow the natural divisions of the material.*" It will be noted that the section headings sometimes are subject-matter headings, sometimes method headings, sometimes fields of application. Hence the material in certain chapters cuts across subject-matter headings, giving the book a "cut up" arrangement. Experiments on learning may be found in the section on Physiological Psychology, on Comparative Psychology, or on Learning and Memory itself. Studies of Perception may be found under

Social Psychology or the Broad Features of Sensory Experience. No clear organization is apparent in the succession of section headings or chapters.

The particular experiments chosen for each of the chapters are somewhat unusual. Although certain of the more typical experiments are represented, there are others that are not commonly included in textbooks. As a result, beginning students using this book are likely to be less well grounded on certain of the more classical experiments.

The author has written the book with the laudable aim of showing that experimentation is not "mere data collection" as the beginning student often thinks. He tries to alert the student to the derivation of principles from experiment by "an extended description of exemplary investigations from the several main divisions of psychology." Readers may take exception to the author's selections both of the main divisions and particular experiments. The diversity of material, lack of a clear organization and wealth of methodological detail may hamper the fulfillment of the author's aim.

CARL PFAFFMAN.

*Brown University.*

BUGELSKI, B. R. *A first course in experimental psychology*. New York: Holt, 1951. \$3.50.

When the student begins laboratory work in psychology, he usually gets his directions from one book and studies experimental psychology from another. Bugelski's book serves both purposes, thereby achieving an integration that is often lacking. The chapter on vision illustrates the plan. Before introducing three experiments, the blind spot, depth perception, and color mixture, the author devotes 13 pages to the stimuli for vision, the size, strength and quality of visual stimuli, color names and wavelengths, and a list of variables in visual experimentation. The author cannot say much about vision in 13 pages, to be sure, but what he says helps the student do his work, gives him some terms, principles and, particularly, some information about experimental methods and their limitations. The student could get along with this book alone, though references to standard sources are mentioned.

There are 27 experiments, including three on psychophysics, five on sensory processes, three labelled motivation, five on learning, two on problem solving, and three labelled emotional behavior. Every one is worth while. According to the introduction, each experiment "has been chosen to teach at least one new technique, control, or procedural feature." The directions seem quite easy to follow; apparently they have been developed and revised in the author's classes.

Most of the experiments can be done with the facilities of small laboratories, but some require more elaborate apparatus. Ingenious construction suggestions and schematic drawings are included, and

Bugelski recommends that the students make their own apparatus. While this plan has certain advantages, it is not as inexpensive as it seems, if the instructor's time and the life of the apparatus are included in the accounting. Furthermore, the students will compare the home-made apparatus they see in psychological laboratories with the standard equipment they see in other laboratories.

The author states in his preface that he makes no attempt to conceal his biases. The sophisticated reader will agree, though the students may not catch on. Bugelski rejects meaningful learning material (p. 324), visceral signs of emotion (p. 368), personality as a whole, (p. 390), ego involvement (pp. 391-393), and so on. He prefers experiments that fit easily into the S-R framework. This book is not intended to encourage students to devise experiments in social or applied psychology. And in the last chapter the author goes out of his way to state "why some problems are neither studied nor solved." Fortunately, most students will skip this chapter, but those who read it will get a strangely limited notion of science. The truth is that scientists have always been solving problems that other scientists told them to avoid. And some students who are introduced to experimental psychology by this book may later find experimental answers for questions that the author tells them not to ask.

The style is chatty and informal. It should be easy to study. But the informality permits the author occasionally to make parenthetical comments that he probably would not make as straightforward sentences. For example, the parenthetical definition of MA on page 74 is actually wrong. And there are many passing digs at psychologists who do not follow the path of modern behaviorism. But these are minor annoyances. Looking at it from the student's point of view and allowing the author the standard number of pet peeves, this book offers an unusually convenient, reasonably accurate introduction to experimental methodology.

DONALD M. JOHNSON.

*Michigan State College.*

MURPHY, GARDNER. *An introduction to psychology*. New York: Harper, 1951. Pp. xvii+583. \$4.25.

The primary motivation of the typical student of beginning psychology is to increase his understanding of people. Sometimes we very effectively smother this original interest, in an urgent compulsion to prove that scientific experiments have been done by psychologists. Sometimes instead we cater to his desires by concocting an innocuous course in mental hygiene or arranging a sort of group-counseling situation. But perhaps there is a more mutually satisfactory solution. Perhaps it is possible to present the facts and principles of psychology in a unified and consistent fashion, but with attention still centered

upon the living, acting *individual*. This is the orientation of Murphy's *Introduction*.

It is not to be inferred that Professor Murphy has striven to reconcile the divergent pedagogical aims through a lukewarm compromise. His general psychology is quite systematic, the logical extension of the viewpoint underlying his 1947 *Personality*, and labelled fairly by Stern's term "personalistic." The very characteristics which make the position meaningful to the systematist are intended to infuse it with vitality for the student. (Unhappily, they may at the same time make the textbook unacceptable to those with differing theoretical predilections.)

It may be said, first of all, that Murphy's chief objective is to describe the actions of the whole person, rather than discuss a succession of discrete topics of traditional psychology. This is in turn reflected both by the major emphases running through the book and by the form of presentation.

1. A recurring theme is the wholeness of the organism, the ceaseless interplay of all parts and functions, the inseparability of those abstracted entities too often segregated in independent chapters. In the beginning, the direction of a cell's development depends upon its relation to other cells; much later, and more complexly, the behavior patterns commonly called "personality" are influenced by the organism's needs and by his perceptual skills, functions themselves of extended growth and spiraling sequences of learning. And so on. If this general principle is not startlingly new, at least it is more faithfully kept before the student here than in the conventional textbook.

2. A complementary emphasis is upon the notion of individuality. Evidence is constantly adduced for the uniqueness, as well as for the continuing flux, of each person's patterns of characteristics.

3. A central place is also given to the socially and culturally determined "self"—the role in which one perceives himself.

By these devices, it is hoped, the reader will be helped to concentrate upon a person's acting in a certain manner rather than upon such cold abstractions as perception or imagination.

The major goal of the book is supported admirably by the organization and style adopted. While there are 30 chapters, with many of the usual titles, there are no really isolated sections, few paragraph headings, no clear breaks in the subject matter. Rather, there is a continuous story about behavior, told almost in essay form. The writer has little preoccupation with detailed descriptions of experiments or formal enumerations of behavioral phenomena. When an observation or principle illuminates the point under consideration, it is stated, rather than being saved to fill its own niche in a fixed outline. Unless an elaboration of the principle is clearly relevant to the developing theme, it is stated quite baldly, with the reference simply noted in one of the 280-odd footnotes.

While this form and content will seem to some instructors to furnish a very satisfactory foundation in psychology, a few criticisms may be



anticipated. Many of these very points, incidentally, will be interpreted by some to be strong points rather than weaknesses.

1. There is too little consideration of the scientific method in general and the experimental paradigm in particular.

2. Specific methodology and findings of psychology are inadequately covered; the writing is less tightly packed with technical material than it might be. The result may be easy reading, but the responsibility for truly representing the field remains unfulfilled.

3. Evidence is often cited too tersely, with the needed controls not at all apparent.

4. An oversimplified picture sometimes results from failure to state contradictory data and theoretical positions.

5. The many literary allusions might falsely be taken to represent the kind of observation acceptable to modern psychology, rather than serving simply as interest-getters and mnemonic devices.

6. There are a few social references so presented as to have almost the faint flavor of propaganda.

Whether these items will balance favorably or unfavorably, one attribute can be commended with confidence. In an effort to soften the technical, scientific tome by "applying psychology," certain elementary textbook writers have resorted to insipid illustrations from an imaginary collegian's life and self-conscious attempts to write in the vernacular. These all too seldom come off, and generally turn out to be only an embarrassment to both teacher and student. Murphy has the happy faculty of writing informally and at the same time unobtrusively, without being sophomoric, trifling, or overly undignified. In this respect, *An Introduction to Psychology* may serve as proof to future writers that one need not be pricked in uncomfortable succession by both horns of the style dilemma. Nor, in fact, by either horn.

FRANK W. FINGER.

*University of Virginia.*

GARRETT, HENRY E. *Psychology*. New York: American Book Co., 1950. Pp. ix+323. \$3.00.

Consistent with his purpose, Professor Garrett has written a book for the general student "who is not planning to major in psychology, and whose acquaintance with academic psychology may begin and end with the introductory course" (p. v.). The fundamental ideas, facts, and principles of traditional psychology are presented and interpreted in terms of the familiar experiences of college students. We recognize the difficult task of such a presentation, avoiding the easy drift toward the static summaries of psychological experiments and theories or to the "social and moral preachments no matter how vital these issues may be to the instructor personally" (p. vi). Perhaps the over-all text would be more attractive and interesting reading if the author had not

remained so religiously close to the classical topics and evidence of general psychology. The lonely one-page treatment of "Tasting and Smelling" contributes little to the main theme of the chapter on "Perception and Observing," which in turn almost ignores the highly stimulating (at least to the student) recent research on perceptual dynamics and phenomenology.

The nine chapter headings illustrate the traditional coverage: What Psychology is About; Heredity and Growth; Motivation and Adjustment; Perception and Observation; Learning: Principles and Methods; Learning and Studying Effectively; Intelligence; Individual Differences and Aptitudes; Personality. Following each chapter are the usual "Questions and topics for discussion." The author has done a superior job in presenting questions which require the student to draw the implications and applications of the principles developed in the chapter.

It is difficult to be briefly systematic and the author chose to curtail discussion of several topics (emotions, sensory processes, thinking, and social factors in personality, etc.). A two-page treatment of emotions, for example, is presented under the heading of "Psychological Motives," and except for the heading, practically no attempt is made to show the systematic importance of emotions in human experience and behavior. Chapters dealing with personality deviations, interpersonal relations, and statistics are purposefully omitted since they are believed to belong more appropriately to specific follow-up courses. Learning, however, is extensively treated in two complete chapters.

The one-course psychology student for whom this book was written will have the opportunity to learn the basic contributions of scientific general psychology. This material is presented in an honest and, we believe, successful attempt to make it meaningful *and* useful. The 323 pages in the text represent the short end of the psychology textbook spectrum. Some of the color we have seen in recent beginning texts has been omitted and it is quite possible that if read in the evening the material would sometimes appear rather gray. We are, however, pleased to note that this well-known author did not sacrifice a "solid" psychology for popular demand.

S. C. ERICKSEN.

*Vanderbilt University.*

PRONKO, N. H., & BOWLES, J. W., JR. *Empirical foundations of psychology*. New York: Rinehart, 1951. Pp. xvi+464. \$3.75.

"The primary motivation toward assembling the works included here," say the authors, "was a sensitivity to writings that they believe presage a revolution in the field of psychology comparable to the Einsteinian one in physics." Presumably, then, the selections included—

some originals of their own, some rewritten from the works of others—may be a harbinger of an “age of psychology.” The chief aim, they assert further, is to “provoke the student to think about fundamental problems in psychology which have often been treated complacently or as closed matters.” The selections are intended to serve as “whetstones against which the student may sharpen his own intellectual tools.”

Swept enthusiastically along by spiralling promises, his appetite further sharpened by assertions that the seasoning of the present psychological dish is attained largely by empirical or operational material (including both man's and nature's experiments), the reader hastens from the hors d'oeuvre of the preface to the pièce de résistance which the authors claim to have assembled.

And what does he find to appease his stimulated hunger? The content of the book, organized under the major topic headings given below, fits the form of the orthodox textbook and in that sense facilitates use of the book as parallel supplementary reading. The sub-topics used to develop the main topics represent careful cullings of unusual material from scattered sources and often provide uncommon insights into common phenomena less well treated in conventional texts. In many instances these sub-topics involve accounts of experiments which in standard texts receive only brief mention (or none at all). For the beginning student it may be that the key to intellectual stimulation comes from such detailed presentation of experimental work.

As an example of selections, the section on Social Behavior contains such items as (a) The social psychological response, (b) Culture, (c) Food: The story of coffee, (d) Dress and fashion: White man's fashions, (e) Superstitions of school children, (f) Voodoo death, (g) Sexual behavior and social level, (h) Cultural factors in gesture, (i) Folie à deux, (j) Contingential responses, (k) Idiosyncratic behavior.

Though reasonably well satisfied with the material presented, the reviewer nevertheless wishes for some rationale to account for the choices made, the omissions, the varying emphasis, and particularly for some insight into the sequence in which the topics are arranged. After the first five sections very little logic appears to govern the order of the remainder. For instance, the phenomena of learning are covered in section 15, the next to the last, while compelling arguments would favor its appearance much earlier as a basis and explanation for much of the material presented before this section but dependent upon it for clarification. The authors may have anticipated this critical comment when they remark in the preface that no attempt had been made to assemble a systematic text and, further, that the aim was “to bring together scattered nuggets from psychological literature.” Nuggets, like pearls on a string, may increase in value if properly related to each other. And the psychologist, like any other scientist, is interested in related facts.

Reaction to the book is determined perhaps less by the material *per se* than by its failure to satisfy the expectations aroused. The heady assertion that the works assembled here "presage a revolution in the field of psychology comparable to the Einsteinian one in physics" can be accepted only if one (a) underestimates Einstein's contributions, and (b) overestimates the efforts of Pronko and Bowles. Few psychologists will wish to do so. The book is a good illustration of the disappointment that can result when an otherwise acceptable product is oversold. The work has merit and deserves an audience—but not a place of honor in the Einsteinian hall of fame.

GEORGE F. J. LEHNER.

*University of California, Los Angeles.*

ARNOLD, WILHELM. *Das Raumerlebnis in Naturwissenschaft und Erkenntnistheorie.* (The experience of space in natural science and epistemology.) Nürnberg: Sebaldus-Verlag, 1949. Pp. 176.

About one-fourth of this monograph is devoted to a historical-theoretical survey of the problem of space perception. This is followed by the description of a new device for testing and measuring acuity of stereoscopic vision. The apparatus is in a way an extension and modification of the one used by Helmholtz in the experiment where three vertical needles are placed in a plane perpendicular to the line of vision. By moving the middle needle out of the plane and towards or away from the observer one is able to determine the threshold values of depth perception. The author's apparatus uses a stereoscope with two sets of three threads in a plane perpendicular to the line of vision. In this device the threads remain in the same plane and perception of depth is achieved through binocular parallax by moving one of the middle threads laterally. It is now also possible to permit the observer to move the middle thread in the other set until all three threads are again perceived as lying in the same plane. The chief advantage of this device lies in the absence of additional cues of depth perception which might be added if a thread is moved closer towards the observer or away from him. Proper measuring devices permit quantitative observations.

The rest of the monograph contains a report of readings obtained at the Psychological Institute of the University of Munich and a discussion of the practical value of the new apparatus. It is suggested that among other things it might be used for training in depth perception and for the determination of individual differences in acuity of depth perception which play a role in certain professions and in piloting a plane.

There are 80 general references to the experience of space taken from the fields of philosophy, psychology, esthetics and physics; and 408 specific references to binocular depth perception.

KARL F. MUENZINGER.

*University of Colorado.*

CATTELL, RAYMOND B. *Personality: A systematic, theoretical and factual study*. New York: McGraw-Hill, 1950. Pp. xii+689. \$5.50.

The aim of this book in the author's words is "to treat personality study by the same scientific standards as are maintained in experimental psychology and to integrate it with research and systematic psychological concepts." This is a very ambitious aim since as the author points out the hypotheses which at present "have a close relationship to facts are restricted in range." It is offered as an antidote to the "rank and wanton pedantry of 'schools' more interested in the social pomp of an all-embracing theory than in a sober desire for truth," a chin-leading statement in a book which itself has systematic theoretical aims.

Starting with the definition of personality as "that which permits a prediction of what a person will do in a given situation" and the truism that scientific advance depends upon exact measurement and exact measurement depends upon exact description, the early chapters deal with varieties of description and measurement. They include a discussion of the devices for obtaining personality data—life history and behavior ratings, self-inventories and tests; a discussion of various techniques of covariation analyses of relationships of persons on tests, tests on persons, and tests on occasions. They include a discussion of surface traits and "source" traits in their various subdivisions or modalities and the statistical determination thereof. In short, the early chapters are a condensation of the author's *Description and Measurement of Personality* (1946). In the reviewer's opinion the condensation often results in a loss of clarity and gives the impression of arbitrariness.

In the remaining chapters an attempt is made to pull together the factorial findings of research literature where validation had been established by three or more independent studies and to point out where further research is needed. Three chapters deal with biological influences in personality under the headings of Inherited, Constitutional Influences, and Psychosomatics. Four chapters are devoted to psycho-dynamics (I. The Intrinsic Structure; II. The Structure of Innate Drives; III. The Structure in the Adjustment Process; IV. The Self—Its Integration, Adaptation and Adjustment). Four chapters are concerned with personality and the social matrix (I and II are on the family: Structure and Setting is the first and Its Relation to Child Personality is the second; III. Group Dynamics; IV. Specific Social Organs). Three chapters concern the abnormal and unadapted personality (I. The School Misfit, the Delinquent, the Criminal; II. The Neurotic; III. The Psychotic). Two chapters are on the life stages in personality (I. Conception to Puberty; II. Adolescence, Maturity and Old Age). A final chapter is devoted to Principles of Personality Formation in which 17 principles of personality are enunciated.

From this sweeping coverage of methods of data collection, of factorial analysis (R, Q, and P), of lists of surface traits derived from clusters of correlations and source traits derived from factoring, constitutional and environmental mold traits in the modalities of ability traits, temperament traits and dynamic traits (goal and sub-goal seeking),



and from the final list of principles of personality formation, this reviewer emerged with no apperceptive coherence or theoretical clarity. The reviewer shares the author's belief that to build substantially in this complex field one must start with empirical descriptive facts, discover relationships, develop constructs which will lead to more precise measurements and higher level organizing constructs, better designed research, and eventually to established principles. Perhaps the data in this field are still too fragmentary, too crude, too unplanfully gathered, and too embedded in semantic imprecision and confusion to permit of systematic coherence, but the reviewer felt that the book did not approximate a return to its original definition of personality as "that which permits a prediction of what a person will do in a given situation." The book, however, should be read by anyone attempting to do serious research in the field of personality, since it contains shrewd and critical evaluations of productive methods of analysis, much important research material, and many suggestions for fruitful research attack.

JEAN WALKER MACFARLANE.

*University of California.*

CATTELL, RAYMOND B. *An introduction to personality study*. New York: Longmans, Green, 1950. \$1.50 (text); \$2.00.

This abridged version of Cattell's recently published text, *Personality: A Systematic, Theoretical and Factual Study* must be viewed in two different lights: first as an introduction to personality study for the sophisticated layman and second as a presentation of the author's conceptualization of personality. Despite Cattell's belief that he has in the main presented a consensus of psychological opinion, he generally states only his own position, and unfortunately he frequently treats alternative positions with open contempt. Phrases such as "we cannot . . . pause to refute every extravagant view that has ever been put forward on human motive" and labels like "extremist" or "more enthusiastic than discreet" are too frequent. As a presentation of Cattell's position, however, the book is clear and readable, logically organized, and well worth reading.

This position can be summarized only briefly. Cattell accepts instincts defined in much the same terms as McDougall used. Several methods for the empirical discovery of instincts are described, but the author clearly favors factor analytic procedures. A basic erg is subsided by various symptomatic attitudes which describe the means by which the need is satisfied. The different attitudes which subside the same erg are expected to have a high correlation which will enable them to be identified by factor analysis. At the same time the basic doctrine that different subsidations of the same erg may substitute for each other and be alternative outlets, is accepted. The existence of ten or twenty basic ergs is reconciled with the psychoanalytic position by means of a postulated maturation process. The psychoanalytic in-

instincts, through differentiation and fusion, mature into the basic ergs of the adult.

The process of adjustment is conceptualized in terms of six adjustment crossroads: (1) stimulated drive, (2) initial frustration, (3) failure of aggression, (4) anxiety and renunciations, (5) the fate of repression, and (6) neurotic reactions. At each crossroad there are various possible sequels, one of which leads to the next crossroad while the others lead back to previous crossroads or to some sort of successful adjustment. This conceptualization is an attempt to clarify the various psychoanalytic mechanisms and the vicissitudes of the instinct. It may not be entirely adequate as a theory, but it is an excellent presentation of a complex set of relationships which is frequently difficult for students to understand.

In terms of this theoretical position, Professor Cattell discusses in a chapter each, abnormal personality, psychosomatics, personality and the culture pattern, and the life course of personality. In two chapters he presents a very readable account of factor analysis and the results of his factor analytic description of personality.

The conceptualization which Cattell builds seems to the reviewer to be rather more a diagrammatic presentation of the points of view which the author has adopted than an integration of these points of view into a single coherent theory. The reconciliation of McDougall, psychoanalysis, and factor analysis is more verbal than real, because the logical difficulties are not faced squarely. For example, the doctrine that different attitudes which subsidiate the same erg are substitutable alternatives for each other implies that in a group of people with the erg at constant strength, the correlation between attitudes subsidiating the same erg would be negative. It is only when the strength of the erg varies that there would be a factor tending toward positive correlation between attitudes subsidiating the same erg. These two opposing tendencies would apparently make factor analysis an insensitive tool for revealing the existence of a common ergic root to a group of attitudes.

Cattell states that the psychoanalytic instincts differentiate and fuse in the course of development to become the basic ergs of the adult. Because the meaning of differentiation and fusion is not clear, the distinction between this process and subsidiation is unclear. Presumably the psychoanalytic theory would demand that the relation between Cattell's ergs and the instinctual drives be one of subsidiation. Thus the crucial issue is left unclear.

In summary, the book makes interesting reading. It is an attempt to do something which badly needs doing, to integrate the various points of view about personality. In view of the heroic proportions of the task it is not surprising that different readers will feel that one or another aspect of the task has been neglected. Nevertheless, to this reader it has achieved only a superficial unity.

ALFRED L. BALDWIN.

*University of Kansas.*

YACORZYNSKI, G. K. *Medical psychology. A basis for psychiatry and clinical psychology.* New York: Ronald Press, 1951. Pp. 535. \$6.00.

In the author's words "The purpose of this book is to present an integrated approach to the understanding of human behavior as a foundation for the study of psychiatry and clinical psychology. Students of medicine require a fundamental background of this kind for their later study of diagnostic methods, psychosomatic medicine, neuroses and psychoses, and finally therapy. Similarly, students majoring in clinical psychology stand in need of an integrated approach oriented primarily toward an understanding of abnormal behavior, without the bias of any particular school."

The book is divided into three parts: "Basic Psychobiological Principles," "Inheritance and Maturation," and "Integration: Structure and Structuralization of Personality." Neither giving the headings of these parts nor the titles of individual chapters conveys either the wealth of material contained in this book or the earnest, and for the most part, successful attempt to integrate theory and fact, psychology and biology. The scope of this book is such that this reviewer has concluded that it could be used as a text in an abnormal psychology course, in an intensive introductory psychology course, and as a good reference for clinical courses. Perhaps its most meritorious aspect is the refreshing, clear and non-superficial manner in which the psychological and biological aspects of functioning are related. The psychology student's perspective upon the problems in his field will be enlarged by the author's biological orientation (and background); similarly, the medical student will get a much-needed orientation toward what is meant by the "total" individual.

SEYMOUR B. SARASON.

*Yale University.*

BARKER, ROGER G., & WRIGHT, HERBERT F. *One boy's day: A specimen record of behavior.* New York: Harper, 1951. Pp. ix+435. \$3.50.

Because of the nature of this publication, it is not possible to review *One Boy's Day* critically. The book contains a record of 14 hours of consecutive observation of a seven-year-old, midwestern, small-town boy, at home, at school, and at play. Each of eight observers recorded in one-minute intervals the objective behavior of Raymond Birch in his natural habitat and interpretive comments regarding that behavior. Thus we see that on April 26, 1949 at 9:42 A.M., Raymond acted in the following manner in his second-grade classroom:

9:42 Raymond hurried out of the music room and back to his classroom, not boisterously, but as though he were eager to return.

"He walked briskly to his desk, on which his May basket materials lay just as he had left them.

"At once he put his whole heart into working on his May basket and getting the paper fastener through the hole at the end of each paper strip.

The theoretical framework within which the observations were collected is not included in this volume. *One Boy's Day* contains, in effect, only raw data. Obviously we cannot evaluate raw data without knowing the frame within which the data were gathered. The reader, therefore, is left with such questions as why these kinds of data and not others, how can the data be used, etc., etc. It is unfortunate, in this reviewer's opinion, that the authors did not enlarge the volume to include their formulation of the meaning and task of psychological ecology.

CELIA BURNS STENDLER.

*University of Illinois.*

## BOOKS AND MATERIALS RECEIVED

BALZ, ALBERT G. A. *Cartesian studies*. New York: Columbia Univ. Press, 1951. Pp. vi+328. \$4.50.

BEKKER, SARAH M., & HARRIS, CATHERINE E. *Research relating to children: An inventory of studies in progress, reported August-December 1950 to the Clearinghouse for Research in Child Life*. Washington, D. C.: Federal Security Agency, Social Security Administration, Children's Bureau, 1950. (Suppl. No. 3.) Pp. 128.

BRAV, STANLEY R. (Ed.) *Marriage and the Jewish tradition: Toward a modern philosophy of family living*. New York: Philosophical Library, 1951. Pp. xiii+218. \$3.75.

BENNETT, WENDELL C. *Area studies in American universities*. New York: Social Science Research Council, 1951. Pp. x+82.

BUROS, OSCAR KRISEN. *Statistical methodology reviews 1941-1950*. New York: John Wiley, 1951. Pp. 457. \$7.00.

CAMERON, NORMAN, & MAGARET, ANN. *Behavior pathology*. Boston: Houghton Mifflin, 1951. Pp. xvi+645. \$5.00.

COMMITTEE ON EDUCATION, TRAINING AND RESEARCH IN RACE RELATIONS OF THE UNIVERSITY OF CHICAGO. *Inventory of research in racial and cultural relations, 1951, 3*, Bull. No. 4.

CORY, D. W. *The homosexual in America: A subjective approach*. New York: Greenberg, 1951. Pp. xvii+326. \$4.00.

DEWEY, RICHARD, & HUMBER, W. J. *The development of human behavior*. New York: Macmillan, 1951. Pp. xiv+672. \$5.50.

EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E., & TYLER R. W. *Intelligence and cultural differences: A study of cultural learning and problem-solving*. Chicago: Univ. of Chicago Press, 1951. Pp. xii+388. \$5.00.

HARMON, FRANCIS L. *Principles of psychology*. (Rev. ed.) Milwaukee, Wis.: Bruce, 1951. Pp. xi+656. \$4.25.

HARROWER, M. R., & STEINER, M. E. *Large scale Rorschach techniques: A manual for the group Rorschach and multiple choice tests*. (2nd ed.) Springfield, Ill.: Charles C Thomas, 1951. Pp. xx+353. \$8.50.

HAYES, CATHY. *The ape in our house*. New York: Harper, 1951. Pp. vi+247. \$3.50.

HOFSTAETTER, PETER R. *Die Psychologie und das Leben*. Wein: Humboldt-Verlag, 1951. Pp. 287.

JAHODA, MARIE, DEUTSCH, MORTON, & COOK, STUART W. *Research methods in social relations: With especial reference to prejudice. Part One: Basic processes. Part Two: Selected techniques*. New York: Dryden



Press, 1951. Vol. I: Pp. x+421, \$3.75. Vol. II: Pp. x+423-759, \$3.75. (\$6.00, set.)

JANIS, IRVING L. *Air war and emotional stress: Psychological studies of bombing and civilian defense*. New York: McGraw-Hill, 1951. Pp. ix+280. \$5.00.

LUDLOW, WILLIAM L. *A syllabus and a bibliography of marriage and the family*. New Concord: Radcliffe Press, 1951. Pp. 309. \$4.00.

MARX, MELVIN H. *Psychological theory: Contemporary readings*. New York: Macmillan, 1951. Pp. xi+585. \$5.00.

MEAD, MARGARET. *Soviet attitudes toward authority*. New York: McGraw-Hill, 1951. Pp. 148. \$4.00.

MELZER, JOHN HENRY. *Functional logic*. Dubuque, Ia.: Wm. C. Brown, 1951. Pp. vii+110. \$3.75.

MILLER, GEORGE A. *Language and communication*. New York: McGraw-Hill. Pp. xiii+298. \$5.00.

OSBORNE, ERNEST G. *The family scrapbook*. New York: Association Press, 1951. Pp. xv+457. \$3.95.

PIÉRON, HENRI, FESSARD, ALFRED, & FRAISSE, PAUL. *L'année psychologique: Cinquantième année, volume jubilaire hommage à Henri Piéron*. Paris: Presses Universitaires de France, 1951. Pp. xvi+718. 2.400 frs.

QUEENER, E. L. *Introduction to social psychology*. New York: William Sloane, 1951. Pp. xiv+493. \$4.25.

SANFORD, CHARLES W., HAND, HAROLD C., & SPALDING, WILLARD B. (Eds.). *The schools and national security: Recommendations for elementary and secondary schools*. (Illinois Secondary School Curriculum Program. Bull. No. 16.) New York: McGraw-Hill, 1951. (Springfield, Ill.: Office of Public Instruction, 1951.) Pp. xiv+292.

SCHUSTER, GEORGE. *Christianity and human relations in industry*. London: Epworth Press, 1951. Pp. 128. 6s, 6d.

SCHWEBEL, MILTON. *The interests of pharmacists*. New York: King's Crown Press, 1951. Pp. xii+84. \$1.75.

SEARS, PAULINE SNEDDEN. *Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence*. *Psychological Monographs* No. 323 (Vol. 65, No. 6). Washington, D. C.: American Psychological Assn., 1951. Pp. iv+42. \$1.50.

STONE, CALVIN P. (Ed.). *Comparative psychology*. (3rd ed.) New York: Prentice-Hall, 1951. Pp. xvii+525. \$6.00.

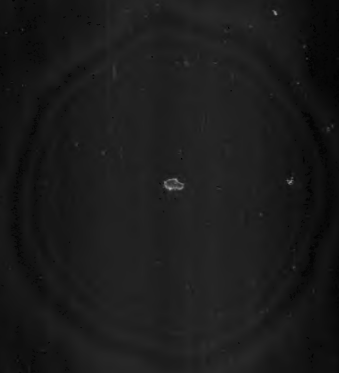
THOMPSON, LAURA. *Personality and government: Findings and recommendations of the Indian Administration Research*. Mexico, D. F.: Ediciones del Instituto Indigenista Interamericano, 1951. (Institute

of Ethnic Affairs, Inc., 810 18th St., N. W., Washington 6, D. C.) Pp. xviii+229. \$2.00.

VALENTINE, C. W. *Psychology: And its bearing on education*. New York: Philosophical Library, 1951. Pp. xvi+674. \$6.00.

ZILBOORG, GREGORY. *Sigmund Freud: His exploration of the mind of man*. New York: Charles Scribner's, 1951. Pp. 132. \$2.00.

C.)  
New  
d of



S

B